NOVEL ODORANT RECEPTORS IN DROSOPHILA

INVENTORS: John R. Carlson, Junhyong Kim, Peter J. Clyne, Coral G. Warr

5 **RELATED APPLICATIONS**

This application claims priority to U.S. provisional patent application Serial No. 60/117,132 filed January 25, 1999 which is herein incorporated by reference in its entirety.

10 **U.S. GOVERNMENT SUPPORT**

FIELD OF THE INVENTION

15 This invention pertains to novel olfactory receptors and to methods of using such receptors. More particularly, this invention pertains to the nucleic acids and amino acids of novel olfactory receptors in *Drosophila* and to methods of using such nucleic acids and amino acids.

20 **BACKGROUND OF THE INVENTION**

Animals can detect a vast array of odors with remarkable sensitivity and discrimination. Olfactory information is first received by olfactory receptor neurons (olfactory receptors), which transmit signals into the central nervous system (CNS) where they are processed, ultimately leading to behavioral responses. An enormous amount of **25** investigation into olfactory function, organization, and development has been carried out in insect model systems for many years (Kaissling *et al.*, (1987) Ann. NY Acad. Sci. 510, 104-112; Hildebrand (1995) Proc. Natl. Acad. Sci. USA 92, 67-74). However, a number of central questions have been refractory to incisive analysis because the receptor

molecules to which odor molecules bind have not been identified, in any insect.

To investigate the molecular mechanisms of olfactory function and development, applicants studied the olfactory system of *Drosophila melanogaster*, which is highly sensitive and capable of odor discrimination (Siddiqi, (1991) Olfaction in *Drosophila*, in:

5 Wysocki & Kare (ed.), Chemical Senses, Marcel Dekker; Carlson (1996) Trends Genet. 12, 175-180). There are two olfactory organs on the adult fly, the third segment of the antenna and the maxillary palp (Figure 1A). In both organs, olfactory receptors are housed in sensory hairs called sensilla. The organization of the approximately 1200 olfactory receptors of the antenna is complex but ordered. On the antenna there are

10 different morphological categories of sensilla: s. trichodea, s. coeloconica, large s. basiconica, and small s. basiconica (Figure 1B). The different morphological categories of sensilla are distributed in overlapping patterns across the surface of the antenna (Figures 1C-F) (Venkatesh & Singh, (1984) Int. J. Insect Morphol. Embryol. 13, 51-63; Stocker, (1994) Roux's Arch. Dev. Biol. 205, 62-72).

15 Electrophysiological studies show that each morphological category of sensilla can be divided into different functional types (denoted by different colors in Figures 1C-F), defined by the characteristic response profiles of their olfactory receptors (Rodrigues *et al.*, (1991) Mol. Gen. Genet. 226, 265-276; Clyne *et al.*, (1997) Invert. Neurosci. 3, 127-135; de Bruyne *et al.*, unpublished results). For s. trichodea, the different functional

20 types are segregated into zones on the surface of the antenna (Figure 1C); segregation is also observed for the different functional types of s. coeloconica (Figure 1D). This zonal organization is less conspicuous for the large and small s. basiconica, of which different functional types are intermingled (Figures 1E-F). Electrophysiological data suggest that there are on the order of thirty different classes of olfactory receptors in the antenna, a

25 rough estimate based upon the odor response profiles of individual olfactory receptors (and in a few cases, the assumption that the neurons of particular functional types of sensilla have unique response profiles).

In contrast to the antenna, the organization of the approximately 120 olfactory

3

receptors of the maxillary palp is less complex. There are approximately 60 s. basiconica on the maxillary palp, each housing two olfactory receptors (Singh & Nayak, (1985) Int. J. Insect Morphol. Embryol. 14, 291-306). The 120 olfactory receptors fall into six different classes based upon their odorant response profiles (Clyne *et al.*, (1999) Neuron

5   22, 339-347; de Bruyne *et al.*, (1999) J. Neurosci. 19, 4520-4532). Neurons of the six ORN classes are always found in characteristic pairs in three functional types of s. basiconica, with the total number of neurons in each class being equal. Each class is distributed broadly over all, or almost all, of the olfactory surface of the maxillary palp.

Thus electrophysiological and anatomical studies suggest that there are on the order

10   of thirty-five classes of olfactory receptors in the adult fly (approximately thirty on the antenna and six on the palp), each class with a distinct odor sensitivity. Classes of olfactory receptors found in the antenna are arrayed in zones, while the classes of olfactory receptors found in the maxillary palp are distributed in a less ordered fashion. olfactory receptors in both the maxillary palp and the antenna extend their axons to the

15   antennal lobe of the brain, where first-order processing of olfactory information occurs. The lobe contains approximately forty olfactory glomeruli, spheroidal modules where ORN axons converge and where their terminal branches form synapses with the dendrites of their target interneurons (Stocker, (1994) Cell Tissue Res. 275, 3-26; Hildebrand & Shepherd, (1997) Annu. Rev. Neurosci. 20, 595-631).

20   One possibility underlying the molecular basis for distinct odor sensitivities for different classes of olfactory receptors is that each class of ORN expresses a unique odorant receptor, as has been proposed for vertebrate olfactory systems (Ngai *et al.*, (1993) Cell 72, 667-680; Ressler *et al.*, (1993) Cell 73, 597-609; Vassar *et al.*, (1993) Cell 74, 309-318; Buck, (1996) Annu. Rev. Neurosci. 19, 517-544; Hildebrand & Shepherd,

25   (1997) Annu. Rev. Neurosci. 20, 595-631). Alternatively, each class of ORN might express a unique combination of a large set of receptors, as found in chemosensory cells of the nematode, *C. elegans* (Troemel *et al.*, (1995) Cell 83, 207-218). Both models call for a family of receptor genes, and several lines of evidence suggest that for insects such a

family would belong to the superfamily of seven-transmembrane G protein-coupled receptors (GPCRs). First, there is evidence that insects generate responses to odorants via GPCR-activated second-messenger systems. For example, a rapid and transient increase in inositol 1,4,5-trisphosphate (IP$_3$) has been observed in response to stimulation with

5   pheromone and other odors using antennal preparations from various insect species (Breer *et al.*, (1990) Nature 345, 65-68; Boekhoff *et al.*, (1993) Insect Biochem. Mol. Biol. 23, 757-762; Wegener *et al.*, (1993) J. Insect Physiol. 39, 153-163). This increase in IP$_3$ can be blocked by pertussis toxin, implicating a G protein signaling cascade (Boekhoff *et al.*, (1990) Cell. Signal. 2, 49-56). In *Drosophila*, norpA mutants, which lack the

10   phospholipase C that is an essential component of phototransduction, also exhibit reduced olfactory responses of the maxillary palp (Riesgo-Escovar *et al.*, (1995) J. Comp. Physiol. A180, 151-160). A second reason to suspect that odorant receptors in *Drosophila* are GPCRs is that GPCRs have been shown to be odorant receptors in both vertebrates and *C. elegans*; moreover, abundant evidence indicates that olfactory information in these other

15   organisms is transduced by GPCR-activated second messenger systems (Buck, (1996) Annu. Rev. Neurosci. 19, 517-544; Bargmann & Kaplan, (1998) Annu. Rev. Neurosci. 21, 279-308). It would thus seem unlikely that a family of receptors that have a completely novel structure and that use a completely different transduction mechanism would have arisen in insects.

20   There have been extensive efforts to identify odorant and pheromone receptors in a variety of insects using a wide range of strategies. These efforts have been driven in part by interest in analyzing receptor genes in the context of highly tractable experimental systems in which there is a wealth of knowledge about olfactory function and organization. For example, *Drosophila* offers the advantages of a model genetic

25   organism together with the ability to measure olfactory function conveniently *in vivo*, through either physiological or behavioral means. Interest in insect odorant receptors has also arisen because of the critical role of olfaction in the attraction of many insect pests to their plant hosts, of insect vectors of disease to their human hosts, and of insects to their

mates. Nevertheless, efforts to identify odorant receptors in insects, based upon searches for genes bearing sequence similarities to odorant receptor genes from other organisms, or on other strategies, have been unsuccessful.

Applicants have discovered a novel multigene family encoding candidate odorant receptors that were identified from the *Drosophila* genomic sequence database. The forty-nine genes described here were discovered using novel computer programs that identify diagnostic features of the protein structure of the seven-transmembrane GPCR superfamily. Members of this new family are highly divergent from previously defined genes. Nearly all of the genes are found to be expressed in one or both of the olfactory organs, and for a number of genes expression is restricted to a subset of olfactory receptors. Applicant's further demonstrate that expression of different genes is initiated at different times during the development of the adult antenna, and that expression of a subset of these candidate receptor genes depends on the POU domain transcription factor, Acj6 (abnormal chemosensory jump 6).

SUMMARY OF THE INVENTION

This invention provides isolated nucleic acid molecules including the following:

a) isolated nucleic acid molecules that encode the amino acid sequences of *Drosophila* Odorant Receptor proteins;

b) isolated nucleic acid molecules that encode protein fragments of at least 6 amino acids of a *Drosophila* Odorant Receptor proteins; and

c) isolated nucleic acid molecules which hybridize to nucleic acid molecules which include nucleotide sequences encoding *Drosophila* Odorant Receptor proteins under conditions of sufficient stringency to produce a clear signal.

This invention also provides such isolated nucleic acid molecules wherein the nucleic acids include at least one exon-intron boundary located in one of the following positions:

a) the nucleotides encoding the amino acids which include the third extracellular

domain of a *Drosophila* Odorant Receptor protein;

     b) the nucleotides encoding the amino acids which include the fourth extracellular domain of a *Drosophila* Odorant Receptor protein; and

     c) the nucleotides encoding the amino acids which include the fourth intracellular domain of a *Drosophila* Odorant Receptor protein.

This invention further provides such isolated nucleic acid molecules which have the nucleic acid sequence of one of the following sequences: SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 and 97.

This invention also provides such isolated nucleic acid molecules operably linked to one or more expression control elements.

This invention further provides vectors which include any of the aforementioned nucleic acid molecules and host cells which include such vectors..

This invention also provides host cells transformed so as to contain any of the aforementioned nucleic acid molecules, wherein such host cells can be either prokaryotic host cells or eukaryotic host cells.

This invention also provides methods for producing proteins or protein fragments wherein the methods include transforming host cells with any of the aforementioned nucleic acids under conditions in which the protein or protein fragment encoded by said nucleic acid molecule is expressed. This invention also provides such methods wherein the host cells are either prokaryotic host cells or eukaryotic host cells. This invention further provides isolated proteins or protein fragments produced by such methods.

This invention provides isolated proteins or protein fragments which include:

     a) isolated proteins encoded by one of the following amino acid sequences: SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98;

     b) isolated protein fragments which include at least 6 amino acids of any of the

- 6 -

following sequences: SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98;

c) isolated proteins which include conservative amino acid substitutions of any of the following sequences: SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98; and

d) naturally occurring amino acid sequence variants of any of the following sequences: SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98.

The present invention further provides such isolated proteins or protein fragments which include at least one of the following conserved amino acids:

a) Leucine in the third extracellular domain of a *Drosophila* Odorant Receptor protein;

b) Histidine in the third extracellular domain of a *Drosophila* Odorant Receptor protein;

c) Cysteine in the sixth transmembrane domain of a *Drosophila* Odorant Receptor protein;

d) Tryptophan in the fourth extracellular domain of a *Drosophila* Odorant Receptor protein;

e) Glutamine in the seventh transmembrane domain of a *Drosophila* Odorant Receptor protein;

f) Proline in the seventh transmembrane domain of a *Drosophila* Odorant Receptor protein;

g) Alanine in the fourth intracellular domain of a *Drosophila* Odorant Receptor protein; and

h) Tyrosine in the fourth intracellular domain of a *Drosophila* Odorant Receptor

-7-

protein.

The present invention also provides isolated antibodies that bind to any of the aforementioned polypeptides.

The present invention also provides such antibodies which are either monoclonal antibodies or polyclonal antibodies.

This invention also provides methods of identifying agents which modulate the expression of any of the aforementioned proteins or protein fragments by:

a) exposing cells which express the proteins or protein fragments to the agents; and

b) determining whether the agent modulates expression of said proteins or protein fragments, thereby identifying agents which modulate the expression of the proteins or protein fragments.

The present invention also provides methods of identifying agents which modulate the activity of any of the aforementioned proteins or protein fragments by:

a) exposing cells which express the proteins or protein fragments to the agents; and

b) determining whether the agents modulate the activity of said proteins or protein fragments, thereby identifying agents which modulate the activity of the proteins or protein fragments.

The present invention also provides such methods where the agent modulates at least one activity of the proteins or protein fragments.

This invention provides methods of identifying agents which modulate the transcription of any of the aforementioned nucleic acid molecules by:

a) exposing cells which transcribe the nucleic acids to the agents; and

b) determining whether the agents modulate transcription of said nucleic acids, thereby identifying agents which modulate the transcription of the nucleic acid.

This invention further provides methods of identifying binding partners for the aforementioned proteins or protein fragments by:

- 8 -

a) exposing said proteins or protein fragments to potential binding partners; and

b) determining if the potential binding partners bind to said proteins or protein fragments, thereby identifying binding partners for the proteins or protein fragments.

The present invention also provides methods of modulating the expression of nucleic acids encoding the aforementioned proteins or protein fragments by administering an effective amount of agents which modulate the expression of the nucleic acids encoding the proteins or protein fragments.

This invention also provides methods of modulating at least one activity of the aforementioned proteins or protein fragments by administering an effective amount of the agents which modulate at least one activity of the proteins or protein fragments.

This invention provides methods of identifying novel olfactory receptor genes by:

a) selecting candidate olfactory receptor genes by screening nucleic acid databases using an algorithm trained to identify seven transmembrane receptors genes;

b) screening said selected candidate olfactory receptor genes by identifying nucleic acid sequences with conserved amino acid residues and intron-exon boundaries common to olfactory receptors, and having open reading frames of sufficient size so as to encode a seven transmembrane receptor; and

c) identifying the novel olfactory receptor genes and measuring the expression of olfactory receptor genes wherein the detection of expression confirms said candidate olfactory genes as olfactory genes.

This invention also provides methods of identifying novel olfactory receptor genes by:

a) selecting candidate olfactory receptor genes by screening nucleic acid databases for nucleic acid sequences with sufficient homology to at least one known olfactory receptor gene;

b) screening said selected candidate olfactory receptor genes by identifying nucleic acids with conserved amino acid residues and intron-exon boundaries common to olfactory receptors, and having open reading frames of sufficient size so as to encode a

- 9 -

seven transmembrane receptor; and

       c) identifying the novel olfactory receptor genes and measuring the expression of olfactory receptor genes wherein the detection of expression confirms said candidate olfactory genes as olfactory genes.

5   The present invention also provides transgenic insects modified to contain any of the aforementioned nucleic acid molecules.

This invention also provides such transgenic insects, wherein the nucleic acid molecules contain mutations that alter expression of the encoded proteins.

BRIEF DESCRIPTION OF THE DRAWINGS

10   Figure 1 An overview of the olfactory system of the adult *Drosophila*. (A) The two olfactory organs of the adult fly, the third antennal segment (arrow) and the maxillary palp (arrowhead), scale bar = 100 μm. (B) Higher magnification of part of a third antennal segment showing the morphological categories of olfactory sensilla: s.

15   basiconica [B], s. trichodea [T] and s. coeloconica [C], scale bar = 5 μm. (C-F) Diagram of the olfactory sensilla on the anterior face of the third antennal segment. The different morphological categories of sensilla are indicated by different shapes, and the colors indicate different functional types of sensilla within each morphological category. Dorsal is at the top and medial is to the left. (C) Distribution of different functional types of s.

20   trichodea. (D) Distribution of different functional types of s. coeloconica. (E) The large s. basiconica are densely clustered in a small dorso-medial region, where the different functional types are intermingled. For simplicity, only two types are shown. (F) The small s. basiconica are widely dispersed, and the different functional types are intermingled.

25

       Figure 2 Genomic organization and hydropathy plots of DOR genes. (A) Genomic organization of DOR genes (not to scale). The genes shown are those identified from 16% of the total genomic sequence; most of the available sequence is from Chromosome

2. The approximate chromosomal location of each gene is indicated. Genes separated by less than one kilobase are jointly underlined. Within each cluster, all genes are oriented in the same direction. The transcriptional orientation of the DOR genes with respect to the chromosome is unknown for 2F.1, 25A.1, 47E.2, 59D.1, and the cluster at 33B. (B) The 2F.1 gene is flanked by two closely linked genes, *fs(1)k10* and *crn*. The arrowheads indicate the 3' end of each gene; for 2F.1 the end of the arrow indicates the position of the polyA+ addition signal sequence. (C) Hydropathy plots of the genes whose expression patterns are shown in Figures 4-6. Hydrophobic peaks predicted by Kyte-Doolittle analysis appear above the center line. The approximate positions of the seven putative transmembrane domains are indicated above the first hydropathy plot.

Figure 3 Amino acid sequence alignment of DOR genes. All DNA sequences were obtained from the BDGP database, and the determination of predicted amino acid sequences is described in the Examples. Residues conserved in >50% of the predicted proteins are shaded. The approximate locations of predicted transmembrane domains 1-7 are indicated. Exon-intron boundaries are shown with vertical lines.

Figure 4 DOR genes are expressed in subsets of olfactory receptor neurons in the maxillary palp. *In situ* hybridizations to tissue sections of maxillary palps. Panel A shows a frontal section; all other sections are sagittal. (A) A 46F.1 probe reveals expression in a subset of olfactory receptors which are broadly distributed. The background staining at the periphery of the organ represents non-specific labeling of the cuticle, observed equally for sense and antisense probes. (B) A 33B.3 probe also hybridizes to a subset of cells. Unlabeled olfactory receptors are visible under the cuticular surface (top center). (C) At higher magnification it can be seen that the cells expressing 46F.1 are neurons. Note the axons projecting from the cells into the nerve (n) which runs through the middle of the maxillary palp. The arrowhead indicates an ORN which is not expressing 46F.1, adjacent to an ORN which is strongly stained. The light

- 11 -

staining of the nerve is background staining, observed equally for sense and antisense probes. (D) 33B.3 is not expressed in the *acj6* null mutant, *acj6⁶*.

Figure 5 DOR genes are expressed in subsets of antennal cells. Shown are *in situ* hybridizations to tissue sections of third antennal segments. In panels A, B, D, and F the plane of section passes through the fluid-filled interior of the antenna. (A,B) A 47E.1 probe hybridizes to a subset of cells which are broadly distributed. (C,D) A 25A.1 probe hybridizes to a smaller subset of cells. The angle of section in panel C differs somewhat from the other panels. (E) A 22A.2 probe hybridizes to a subset of cells in the dorso-medial region where the large s. basiconica are located. (F) 22A.2 is expressed in the *acj6⁶* mutant, in contrast to 33B.3 (Figure 4D). (G) Summary of distributions of labeled cells for 47E.1 (open circles), 25A.1 (black dots), and 22A.2 (gray dots) on the anterior face of the antenna, based on analysis of expression in 30-50 antennae for each gene.

Figure 6 Expression of DOR genes during antennal development. *In situ* hybridizations to tissue sections of third antennal segments at different times during pupal development. The times indicated refer to hours APF (after puparium formation). Arrows indicate labeled cells. (A) Expression of 22A.2 is not observed at 54 hours APF. Note that background staining is absent in sections taken at 54 hours (or at earlier times), presumably due to the immaturity of the cuticle. (B) Expression of 22A.2 is observed at 60 hours APF. (C) 47E.1 expression is not observed at 72 hours APF. Background staining is observed with both sense and antisense probes on the cuticular surface of the sacculus (s), a multi-chambered sensory pit and the dot at the bottom of the third antennal segment is non-specific staining of a section of tracheal tissue. (D) Expression of 47E.1 is detected at 93 hours APF. (E) The odor binding protein OS-E is not expressed at 72 hours APF. The small dots at the bottom of the antenna are non-specific staining of a section of tracheal tissue, observed with both sense and antisense probes. (F) Abundant

- 12 -

13

expression of OS-E is seen at 93 hours APF.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

I. Specific Embodiments

5   ### A. *Drosophila* Olfactory Receptor Proteins

The present invention provides a family of isolated proteins, allelic variants of the proteins, and conservative amino acid substitutions of the proteins. As used herein, protein or polypeptide refers to any one of the proteins that has the amino acid sequence depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36,

10   38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98. The invention also includes naturally occurring allelic variants and proteins that have a slightly different amino acid sequence than that specifically recited above. Allelic variants, though possessing a slightly different amino acid sequence than those recited above, will still have the same or similar biological

15   functions associated with any of the amino acid proteins.

As used herein, the family of proteins related to any one of the amino acid sequences depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98 refers to proteins that have been isolated from organisms in

20   addition to *Drosophila*. The methods used to identify and isolate other members of the family of proteins related to these amino acid proteins are described below.

The proteins of the present invention are preferably in isolated form. As used herein, a protein is said to be isolated when physical, mechanical or chemical methods are employed to remove the protein from cellular constituents that are normally associated

25   with the protein. A skilled artisan can readily employ standard purification methods to obtain an isolated protein.

The proteins of the present invention further include conservative amino acid substitution variants (*i.e.*, conservative) of the proteins herein described. As used herein,

a conservative variant refers to at least one alteration in the amino acid sequence that does not adversely affect the biological functions of the protein. A substitution, insertion or deletion is said to adversely affect the protein when the altered sequence prevents or disrupts a biological function associated with the protein. For example, the overall

5   charge, structure or hydrophobic-hydrophilic properties of the protein can be altered without adversely affecting a biological activity. Accordingly, the amino acid sequence can often be altered, for example to render the peptide more hydrophobic or hydrophilic, without adversely affecting the biological activities of the protein.

Ordinarily, the allelic variants, the conservative substitution variants, and the

10   members of the protein family, will have an amino acid sequence having at least 30% amino acid sequence identity with the sequences set forth in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98 more preferably at least 35%, even more preferably at least 40%, and most preferably at least

15   45%. Identity or homology with respect to such sequences is defined herein as the percentage of amino acid residues in the candidate sequence that are identical with the known peptides, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent homology, and not considering any conservative substitutions as part of the sequence identity. N-terminal, C-terminal or internal

20   extensions, deletions, or insertions into the peptide sequence shall not be construed as affecting homology.

In addition to amino acid sequence identity, the proteins of the present invention have seven transmembrane domains as defined by hydropathy analysis (Kyte & Doolittle, (1982) J. Mol. Biol. 157, 105-132). Furthermore, the proteins of the present invention

25   have conserved amino acid residues in defined domains of the protein. For example, the proteins of the present invention have at least one of the following conserved amino acids as depicted in Figure 3, including but not limited to, Leucine in the third extracellular domain; Histidine in the third extracellular domain; Cysteine in the sixth transmembrane

- 14 -

15

domain; Tryptophan in the fourth extracellular domain; Glutamine in the seventh transmembrane domain; Proline in the seventh transmembrane domain; Alanine in the fourth intracellular domain; or Tyrosine in the fourth intracellular domain. In addition, the conserved amino acids may be selected from any of the amino acid residues indicated

5   as being conserved among DOR proteins as depicted in Figure 3 (shaded).

Thus, the proteins of the present invention include molecules having the amino acid sequence disclosed in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98; fragments thereof having a consecutive

10   sequence of at least about 3, 4, 5, 6, 10, 15, 20, 25, 30, 35 or more amino acid residues of the proteins, for instance, antigenic fragments such as those found in the extracellular domains of the protein (see Figure 3); amino acid sequence variants wherein an amino acid residue has been inserted N- or C-terminal to, or within, the disclosed sequence; and amino acid sequence variants of the disclosed sequences, or their fragments as defined

15   above, that have been substituted by another residue. Contemplated variants further include those containing predetermined mutations by, *e.g.*, homologous recombination, site-directed or PCR mutagenesis, and the corresponding proteins of other insect species, including but not limited to the order *Diptera, Lepidoptera, Homopterera* and *Coleoptera*, within these orders, preferably the genus *Drosophila, Anopheles, Aedes, Ceratitis,*

20   *Muscidae, Culicidae, Anagasta* and *Popilla* and the alleles or other naturally occurring variants of the family of proteins; and derivatives wherein the protein has been covalently modified by substitution, chemical, enzymatic, or other appropriate means with a moiety other than a naturally occurring amino acid (for example a detectable moiety such as an enzyme or radioisotope).

25   As described below, members of the family of proteins can be used: 1) to identify agents which modulate at least one activity of the protein; 2) to identify binding partners for the protein, 3) as an antigen to raise polyclonal or monoclonal antibodies, and 4) in methods to modify insect behavior.

- 15 -

16

B. Nucleic Acid Molecules

The present invention further provides nucleic acid molecules which encode any of the proteins having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98 and the related proteins herein described, preferably in isolated form. As used herein, "nucleic acid" is defined as RNA or DNA that encodes a protein or peptide as defined above, is complementary to a nucleic acid sequence encoding such peptides, hybridizes to such a nucleic acid and remains stably bound to it under appropriate stringency conditions, or encodes a polypeptide sharing at least 75% sequence identity, preferably at least 80%, and more preferably at least 85%, with the peptide sequences in conserved domains. Specifically contemplated are genomic DNA, cDNA, mRNA and antisense molecules, as well as nucleic acids based on alternative backbones or including alternative bases whether derived from natural sources or synthesized. Such hybridizing or complementary nucleic acids, however, are defined further as being novel and non-obvious over any prior art nucleic acid including that which encodes, hybridizes under appropriate stringency conditions, or is complementary to nucleic acid encoding a protein according to the present invention.

Homology or identity at the amino acid or nucleotide level is determined by **BLAST** (Basic Local Alignment Search Tool) analysis using the algorithm employed by the programs **blastp, blastn, blastx, tblastn** and **tblastx** (Karlin *et al.*, (1990) Proc. Natl. Acad. Sci. USA 87, 2264-2268 and Altschul, (1993) J. Mol. Evol. 36, 290-300, fully incorporated by reference) which are tailored for sequence similarity searching. The approach used by the **BLAST** program is to first consider similar segments between a query sequence and a database sequence, then to evaluate the statistical significance of all matches that are identified and finally to summarize only those matches which satisfy a preselected threshold of significance. For a discussion of basic issues in similarity searching of sequence databases (see Altschul *et al.*, (1994) Nature Genetics 6, 119-129

- 16 -

17

which is fully incorporated by reference). The search parameters for **histogram,**
descriptions, alignments, expect (*i.e.,* the statistical significance threshold for reporting
matches against database sequences), **cutoff, matrix** and **filter** are at the default settings.
The default scoring matrix used by **blastp, blastx, tblastn,** and **tblastx** is the

5   **BLOSUM62** matrix (Henikoff *et al.,* (1992) Proc. Natl. Acad. Sci. USA 89,
10915-10919, fully incorporated by reference). For **blastn,** the scoring matrix is set by
the ratios of **M** (*i.e.,* the reward score for a pair of matching residues) to **N** (*i.e.,* the
penalty score for mismatching residues), wherein the default values for **M** and **N** are 5
and -4, respectively.

10   "Stringent conditions" are those that (1) employ low ionic strength and high
temperature for washing, for example, 0.5 M sodium phosphate buffer at pH 7.2, 1 mM
EDTA at pH 8.0 in 7% SDS at either 65°C or 55°C, or (2) employ during hybridization a
denaturing agent such as formamide, for example, 50% formamide with 0.1% bovine
serum albumin, 0.1% Ficoll, 0.1% polyvinylpyrrolidone, 0.05 M sodium phosphate buffer

15   at pH 6.5 with 0.75 M NaCl, 0.075 M sodium citrate at 42°C. Another example is use of
50% formamide, 5× SSC (0.75 M NaCl, 0.075 M sodium citrate), 50 mM sodium
phosphate at pH 6.8, 0.1% sodium pyrophosphate, 5× Denhardt's solution, sonicated
salmon sperm DNA (50 µg/ml), 0.1% SDS and 10% dextran sulfate at 55°C, with washes
at 55°C in 0.2× SSC and 0.1% SDS. A skilled artisan can readily determine and vary the

20   stringency conditions appropriately to obtain a clear and detectable hybridization signal.
Preferred molecules are those that hybridize under the above conditions to the
complements of SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33,
35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81,
83, 85, 87, 89, 91, 93, 95 and 97, and which encode a functional protein.

25   As used herein, a nucleic acid molecule is said to be "isolated" when the nucleic
acid molecule is substantially separated from contaminant nucleic acid encoding other
polypeptides from the source of nucleic acid.

     The present invention further provides fragments of any one of the encoding nucleic

acids molecules. As used herein, a fragment of an encoding nucleic acid molecule refers to a small portion of the entire protein coding sequence. The size of the fragment will be determined by the intended use. For example, if the fragment is chosen so as to encode an active portion of the protein, the fragment will need to be large enough to encode the

5   functional region(s) of the protein. For instance, fragments of the invention encode antigenic fragments such as the extracellular loops or N-terminal domain of the protein depicted in SEQ ID NO: 2 and as set forth in Figure 3. If the fragment is to be used as a nucleic acid probe or PCR primer, then the fragment length is chosen so as to obtain a relatively small number of false positives during probing and priming.

10   Fragments of the encoding nucleic acid molecules of the present invention (*i.e.*, synthetic oligonucleotides) that are used as probes or specific primers for the polymerase chain reaction (PCR), or to synthesize gene sequences encoding proteins of the invention can easily be synthesized by chemical techniques, for example, the phosphotriester method of Matteucci *et al.*, (1981) J. Am. Chem. Soc. 103, 3185-3191) or using automated synthesis

15   methods. In addition, larger DNA segments can readily be prepared by well known methods, such as synthesis of a group of oligonucleotides that define various modular segments of the gene, followed by ligation of oligonucleotides to build the complete modified gene.

       The encoding nucleic acid molecules of the present invention may further be modified so as to contain a detectable label for diagnostic and probe purposes. A variety

20   of such labels are known in the art and can readily be employed with the encoding molecules herein described. Suitable labels include, but are not limited to, fluorescent-labeled, biotin-labeled, radio-labeled nucleotides and the like. A skilled artisan can employ any of the art known labels to obtain a labeled encoding nucleic acid molecule.

       Modifications to the primary structure itself by deletion, addition, or alteration of the

25   amino acids incorporated into the protein sequence during translation can be made without destroying the activity of the protein. Such substitutions or other alterations result in proteins having an amino acid sequence encoded by a nucleic acid falling within the contemplated scope of the present invention.

- 18 -

C. Isolation of Other Related Nucleic Acid Molecules

As described above, the identification and characterization of the nucleic acid molecules having SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35,

5   37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 and 97 allows a skilled artisan to isolate nucleic acid molecules that encode other members of the protein family in addition to the sequences herein described. Further, the presently disclosed nucleic acid molecules allow a skilled artisan to isolate nucleic acid molecules that encode other members of the family of proteins in addition to the protein

10   having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98.

Essentially, a skilled artisan can readily use any one of the amino acid sequences selected from SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36,

15   38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98, to generate antibody probes to screen expression libraries prepared from appropriate cells. Typically, polyclonal antiserum from mammals such as rabbits immunized with the purified protein (as described below) or monoclonal antibodies can be used to probe a cDNA or genomic expression library to obtain the appropriate coding

20   sequence for other members of the protein family. The cloned cDNA sequence can be expressed as a fusion protein, expressed directly using its own control sequences, or expressed by constructions using control sequences appropriate to the particular host used for expression of the enzyme.

Alternatively, a portion of the coding sequence herein described can be synthesized

25   and used as a probe to retrieve DNA encoding a member of the protein family from any organism. Oligomers containing approximately 18-20 nucleotides (encoding about a six to seven amino acid stretch) are prepared and used to screen genomic DNA or cDNA libraries to obtain hybridization under stringent conditions or conditions of sufficient stringency to

- 19 -

70

eliminate an undue level of false positives.

Additionally, pairs of oligonucleotide primers can be prepared for use in a polymerase chain reaction (PCR) to selectively clone an encoding nucleic acid molecule. A PCR denature/anneal/extend cycle for using such PCR primers is well known in the art and can

5   readily be adapted for use in isolating other encoding nucleic acid molecules. For example, degenerate primers can be used to clone any DOR gene across species. Specifically, based on the sequence information derived from the family of DORs, degenerate primers can be designed based on conserved sequences among olfactory receptors, which can then be used to clone nucleic acid molecules encoding olfactory receptor proteins from other species of

10   insects.

Applicants have also identified a method for isolating nucleic acid molecules that encode other members of the protein family in addition to the sequences herein described. Essentially, a two-step strategy is employed to identify odorant receptor genes from the genomic database. First, a computer algorithm was designed to search genomic sequences for

15   open reading frames (ORFs) from candidate odorant receptor genes. Second, RT-PCR is used to determine if transcripts from any of these ORFs are expressed in olfactory organs.

The algorithm is used to identify GPCR genes using statistical characterization of amino acid physico-chemical profiles in combination with a non-parametric discriminant function. The algorithm is trained on a set of putative sequences from a database. In the first

20   step, three sets of descriptors are used to summarize the physico-chemical profiles of the sequences. These are GES scale of hydropathy (Engelman *et al.*, (1986) Annu. Rev. Biophys. Biophys. Chem. 15, 321-353), polarity (Brown, (1991) Molecular Biology Labfax, Academic Press), and amino acid usage frequency. For the first two of these measurements, a computed sliding window profile is employed (White, (1994) Membrane Protein Structure, Oxford

25   University Press) using a kernel of a certain number of amino acids as a constant function convoluted with a certain number of amino acids as a Gaussian function. These profiles are then summarized with three statistics; the periodicity, average derivative and the variance of the derivative.

- 20 -

Each sequence is then characterized by multiple variables using a non-parametric linear discriminant function that is optimized to separate the known family proteins from random proteins in the training set. The same linear discriminant function with the scores derived from the training set is used to screen any nucleic acid database for candidate genes.

5 The candidate sequences are given significance values by an odds ratio of the proteins and non-family proteins, computed using the observed empirical distribution of the training set. Those sequences with a sufficiently high odds ratio are considered for further analysis. The algorithm can also be used to identify any protein family by altering the training set of sequences.

10 The method of identification further includes steps for identifying novel olfactory receptor genes comprising selecting candidate olfactory receptor genes by screening a nucleic acid database using an algorithm trained to identify seven transmembrane receptors genes; screening said selected candidate olfactory receptor genes by identifying nucleic acid sequences with conserved amino acid residues and intron-exon boundaries common to

15 olfactory receptors, and open reading frames of sufficient size as to encode a seven transmembrane receptor. As an additional step, the expression of olfactory receptor genes is measured to confirm candidate olfactory gene as an olfactory gene. The exon-intron boundaries and conserved amino acid residues may be selected from any of the positions depicted in Figure 3. Alternatively, selecting candidate olfactory receptor genes by screening

20 a nucleic acid database for nucleic acid sequences with sufficient homology to at least one known olfactory receptor gene is also encompassed in the invention. In a preferred embodiment, the nucleic acid database is a genomic database, an EST database or even an olfactory receptor database as previously described (Skoufos *et al.*, (1999) Nucleic Acids Research 27, 343-345).

25 In one example of the invention, the training set could consist of a subset of seven transmembrane proteins such as dopaminergic receptors and could be used to search genomic sequences for new subtypes of dopaminergic receptors. In another example, the training set could consist of ion channels and could be used to identify new subtypes of ion channels in a

- 21 -

particular family. In yet another example, the training set could consist of known sequences coding for a receptors from a particular family and could be used to identify homologs across species. Specifically, olfactory receptors of one species could be used as a training set to identify olfactory receptors in another species.

5

D. rDNA molecules containing a DNA molecule

The present invention further provides recombinant DNA molecules (rDNAs) that contain a coding sequence. As used herein, a rDNA molecule is a DNA molecule that has been subjected to molecular manipulation *in situ*. Methods for generating rDNA molecules are well known in the art, for example, see Sambrook *et al.*, (1985) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory Press. In the preferred rDNA molecules, a coding DNA sequence is operably linked to expression control sequences or vector sequences.

The choice of vector and expression control sequences to which one of the protein family encoding sequences of the present invention is operably linked depends directly, as is well known in the art, on the functional properties desired, *e.g.*, protein expression, and the host cell to be transformed. A vector contemplated by the present invention is at least capable of directing the replication or insertion into the host chromosome, and preferably also expression, of the structural gene included in the rDNA molecule.

Expression control elements that are used for regulating the expression of an operably linked protein encoding sequence are known in the art and include, but are not limited to, inducible promoters, constitutive promoters, secretion signals, and other regulatory elements. Preferably, the inducible promoter is readily controlled, such as being responsive to a nutrient in the host cell's medium.

In one embodiment, the vector containing a coding nucleic acid molecule will include a prokaryotic replicon, *i.e.*, a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extra-chromosomally in a prokaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are

well known in the art. In addition, vectors that include a prokaryotic replicon may also include a gene whose expression confers a detectable marker such as a drug resistance. Typical bacterial drug resistance genes are those that confer resistance to ampicillin or tetracycline.

5   Vectors that include a prokaryotic replicon can further include a prokaryotic or bacteriophage promoter capable of directing the expression (transcription and translation) of the coding gene sequences in a bacterial host cell, such as *E. coli*. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial hosts

10   are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention. Typical of such vector plasmids are pUC8, pUC9, pBR322 and pBR329 available from BioRad Laboratories, pPL and pKK223 available from Pharmacia.

Expression vectors compatible with eukaryotic cells, preferably those compatible with

15   vertebrate cells such as insect cells, can also be used to form a rDNA molecules that contains a coding sequence. Eukaryotic cell expression vectors are well known in the art and are available from several commercial sources. Typically, such vectors are provided containing convenient restriction sites for insertion of the desired DNA segment. Typical of such vectors are pSVL and pKSV-10 (Pharmacia), pBPV-1/pML2d (International Biotechnologies, Inc.),

20   pTDT1 (ATCC, #31255), the vector pCDM8 described herein, and the like eukaryotic expression vectors. Vectors may be modified to include insect cell specific promoters if needed.

Eukaryotic cell expression vectors used to construct the rDNA molecules of the present invention may further include a selectable marker that is effective in an eukaryotic

25   cell, preferably a drug resistance selection marker. A preferred drug resistance marker is the gene whose expression results in neomycin resistance, *i.e.*, the neomycin phosphotransferase (*neo*) gene (Southern *et al.*, (1982) J. Mol. Appl. Genet. 1, 327-341). Alternatively, the selectable marker can be present on a separate plasmid, and the two vectors are introduced by

co-transfection of the host cell, and selected by culturing in the appropriate drug for the selectable marker.

E. Host Cells Containing an Exogenously Supplied Coding Nucleic Acid

5   The present invention further provides host cells transformed with a nucleic acid molecule that encodes a protein of the present invention. The host cell can be either prokaryotic or eukaryotic. Eukaryotic cells useful for expression of a protein of the invention are not limited, so long as the cell line is compatible with cell culture methods and compatible with the propagation of the expression vector and expression of the gene product. Preferred

10   eukaryotic host cells include, but are not limited to, yeast, insect and mammalian cells, preferably insect cells such as those from a *Drosophila* cell line. Preferred *Drosophila* host cells include *Drosophila* Schneider line 2, and the like insect tissue culture cell lines.

Any prokaryotic host can be used to express a rDNA molecule encoding a protein of the invention. The preferred prokaryotic host is *E. coli*.

15   Transformation of appropriate cell hosts with a rDNA molecule of the present invention is accomplished by well known methods that typically depend on the type of vector used and host system employed. With regard to transformation of prokaryotic host cells, electroporation and salt treatment methods are typically employed, see, for example, Cohen *et al.*, (1972) Proc. Natl. Acad. Sci. USA 69, 2110-2114; and Maniatis *et al.*, (1982) Molecular

20   Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory Press. With regard to transformation of vertebrate cells with vectors containing rDNAs, electroporation, cationic lipid or salt treatment methods are typically employed, see, for example, Graham *et al.*, (1973) Virology 52, 456-467; and Wigler *et al.*, (1979) Proc. Natl. Acad. Sci. USA 76, 1373-1376.

Successfully transformed cells, *i.e.*, cells that contain a rDNA molecule of the present

25   invention, can be identified by well known techniques including the selection for a selectable marker. For example, cells resulting from the introduction of an rDNA of the present invention can be cloned to produce single colonies. Cells from those colonies can be harvested, lysed and their DNA content examined for the presence of the rDNA using a

25

method such as that described by Southern, (1975) J. Mol. Biol. 98, 503-517; or Berent *et al.*, (1985) Biotech. Histochem. 3, 208; or the proteins produced from the cell assayed via an immunological method.

F. Production of Recombinant Proteins using a rDNA Molecule

The present invention further provides methods for producing a protein of the invention using nucleic acid molecules herein described. In general terms, the production of a recombinant form of a protein typically involves the following steps: First, a nucleic acid molecule is obtained that encodes a protein of the invention, such as any of the nucleic acid molecule depicted in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 and 97. The nucleic acid molecule is then preferably placed in operable linkage with suitable control sequences, as described above, to form an expression unit containing the protein open reading frame. The expression unit is used to transform a suitable host and the transformed host is cultured under conditions that allow the production of the recombinant protein. Optionally the recombinant protein is isolated from the medium or from the cells; recovery and purification of the protein may not be necessary in some instances where some impurities may be tolerated.

Each of the foregoing steps can be done in a variety of ways. For example, the desired coding sequences may be obtained from genomic fragments and used directly in appropriate hosts. The construction of expression vectors that are operable in a variety of hosts is accomplished using appropriate replicons and control sequences, as set forth above. The control sequences, expression vectors, and transformation methods are dependent on the type of host cell used to express the gene and were discussed in detail earlier. Suitable restriction sites can, if not normally available, be added to the ends of the coding sequence so as to provide an excisable gene to insert into these vectors. A skilled artisan can readily adapt any host-expression system known in the art for use with the nucleic acid molecules of the invention to produce recombinant protein.

- 25 -

26

G. Methods to Identify Binding Partners

Another embodiment of the present invention provides methods for use in isolating and identifying binding partners of any of the DOR proteins of the invention. In detail, a

5 protein of the invention is mixed with a potential binding partner or an extract or fraction of a cell under conditions that allow the association of potential binding partners with the protein of the invention. After mixing, peptides, polypeptides, proteins or other molecules that have become associated with a protein of the invention are separated from the mixture. The binding partner that bound to the protein of the invention can then be removed and further

10 analyzed. To identify and isolate a binding partner, the entire protein, for instance a protein comprising the entire amino acid sequence of any of the proteins depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98 can be used. Alternatively, a fragment of any of the proteins can be used.

15 As used herein, a cellular extract refers to a preparation or fraction which is made from a lysed or disrupted cell. The preferred source of cellular extracts will be cells derived from *Drosophila*, for instance, antennae and maxillary palp cellular extract.

A variety of methods can be used to obtain an extract of a cell. Cells can be disrupted using either physical or chemical disruption methods. Examples of physical disruption

20 methods include, but are not limited to, sonication and mechanical shearing. Examples of chemical lysis methods include, but are not limited to, detergent lysis and enzyme lysis. A skilled artisan can readily adapt methods for preparing cellular extracts in order to obtain extracts for use in the present methods.

Once an extract of a cell is prepared, the extract is mixed with any of the proteins of

25 the invention under conditions in which association of the protein with the binding partner can occur. A variety of conditions can be used, the most preferred being conditions that closely resemble conditions found in the cytoplasm of a *Drosophila* cell. Features such as osmolarity, pH, temperature, and the concentration of cellular extract used, can be varied to optimize the

association of the protein with the binding partner.

After mixing under appropriate conditions, the bound complex is separated from the mixture. A variety of techniques can be utilized to separate the mixture. For example, antibodies specific to a protein of the invention can be used to immunoprecipitate the binding

5   partner complex. Alternatively, standard chemical separation techniques such as chromatography and density-sediment centrifugation can be used.

After removal of non-associated cellular constituents found in the extract, the binding partner can be dissociated from the complex using conventional methods. For example, dissociation can be accomplished by altering the salt concentration or pH of the mixture.

10   To aid in separating associated binding partner pairs from the mixed extract, the protein of the invention can be immobilized on a solid support. For example, the protein can be attached to a nitrocellulose matrix or acrylic beads. Attachment of the protein to a solid support aids in separating peptide-binding partner pairs from other constituents found in the extract. The identified binding partners can be either a single protein or a complex made up of

15   two or more proteins. Alternatively, binding partners may be identified using a Far-Western assay according to the procedures of Takayama et al., (1997) Methods Mol. Biol. 69, 171-184 or identified through the use of epitope tagged proteins or GST fusion proteins.

Alternatively, the nucleic acid molecules of the invention can be used in a yeast two-hybrid system. The yeast two-hybrid system has been used to identify other protein partner

20   pairs (Alifragis et al., (1997) Proc. Natl. Acad. Sci. USA 94, 13099-13104; Dong et al., (1999) Gene 237, 421-428) and can readily be adapted to employ the nucleic acid molecules herein described.

In another embodiment, binding partners may be identified in insects using single unit recordings as previously described (Kaissling, (1995) Single unit and electroantennogram

25   recordings in insect olfactory organs, in: Spielman & Brand (ed.) Experimental Cell Biology of Taste and Olfaction, CRC Press). Using single unit recordings in vivo, response profiles are established for potential ligands, these profiles are then categorized into distinct functional classes indicative of distinct receptor-ligand interactions (see, e.g., U.S. Patent No. 5,993,778).

- 27 -

Single unit recordings in transgenic insects which contain transgenes resulting in over- or under-expression of a gene are also useful for identifying and characterizing ligands which bind to multiple olfactory receptors as well as identifying characterizing new olfactory receptors.

5   The nucleic acids of the invention and their corresponding proteins can be used on an array or microarray for high-throughput screening for agents which interact with either the nucleic acids of the invention or their corresponding proteins. An "array" or "microarray" generally refers to a grid system which has each position or probe cell occupied by a defined nucleic acid fragments also known as oligonucleotides. The arrays themselves are sometimes

10   referred to as "chips" or "biochips". High-density nucleic acid and protein microarrays often have thousands of probe cells in a variety of grid styles.

        A typical molecular detection chip includes a substrate on which an array of recognition sites, binding sites or hybridization sites are arranged. Each site has a respective molecular receptor which binds or hybridizes with a molecule having a predetermined

15   structure. The solid support substrates which can be used to form surface of the array or chip include organic and inorganic substrates, such as glass, polystyrenes, polyimides, silicon dioxide and silicon nitride. For direct attachment of probes to the electrodes, the electrode surface must be fabricated with materials capable of forming conjugates with the probes.

        Once the array is fabricated, a sample solution is applied to the molecular detection

20   chip and molecules in the sample bind or hybridize at one or more sites. The sites at which binding occurs are detected, and one or more molecular structures within the sample are subsequently deduced. Detection of labeled batches is a traditional detection strategy and includes radioisotope, fluorescent and biotin labels, but other options are available, including electronic signal transduction.

25   Polymer arrays of nucleic acid probes can be used to extract information from, for example, nucleic acid samples. These samples are exposed to the probes under conditions that permit binding. The arrays are then scanned to determine to which probes the sample molecules have interacted with the nucleic acids of the polymer array. One can obtain

29

information by careful probe selection and using algorithms to compare patterns of interactions. For example, the method is useful in screening for novel olfactory receptors in multiple organisms. For example, *Drosophila* degenerate olfactory receptor oligonucleotide arrays can be used to examine a nucleic acid sample from another insect species in order to

5   identify novel olfactory receptors in that species.

In typical applications, a complex solution containing one or more substances to be characterized contacts a polymer array comprising nucleic acids. For example, the array is comprised of nucleic acid probes. The probes of the array can be either DNA or RNA, which may be either single-stranded or double-stranded. In a preferred embodiment of the

10   invention, the probes are arranged (either by immobilization, typically by covalent attachment, of a pre-synthesized probe or by synthesis of the probe on the substrate) on the substrate or chips in lanes stretching across the chip and separated, and these lanes are in turned arranged in blocks of preferably five lanes, although blocks of other sizes will have useful application. The present invention provides individual probes, sets of probes, and

15   arrays of probe sets on chips, in specific patterns which are used to characterize the substances in a complex mixture by producing a distinct image which is representative of the binding interactions between the probes on the chip and the substances in the complex mixture. The pattern of hybridization to the chip allows inferences to be drawn about the substances present in the complex mixture.

20   The substances in the complex solution will bind to the nucleic acids on the array. The substances of the complex mixture which bind to the nucleic acids of the array may include, but are not limited to, complementary nucleic acids, non-complementary nucleic acids, proteins, antibodies, oligosaccharides, etc. The types of binding may include, but are not limited to, specific and non-specific, competitive and non-competitive, allosteric,

25   cooperative, non-cooperative, complementary and non-complementary, etc. For example, the nucleic acids of the array can bind to complementary nucleic acids in the complex mixture but can also bind in a tertiary manner, independent of base pairing, to non-complementary nucleic acids.

- 29 -

30

The nucleic acids of the array or the substances of the complex mixture may be tagged with a detectable label. The detectable label can be, for example, a luminescent label, a light scattering label or a radioactive label. Accordingly, locations at which substances interact can be identified by either determining if the signal of the label has been quenched by binding or

5   identifying locations where the signal of the label is present in cases where the substances of the complex mixture have been labeled. Based on the locations where binding is detected, information regarding the complex mixture can be obtained.

The methods of this invention will find particular use wherever high through-put of samples is required. In particular, this invention is useful in ligand screening settings and for

10   determining the composition of complex mixtures.

Polypeptides are an exemplary system for exploring the relationship between structure and function in biology. When the twenty naturally occurring amino acids are condensed into a polymeric molecule they form a wide variety of three-dimensional configurations, each resulting from a particular amino acid sequence and solvent condition. For example, the

15   number of possible polypeptide configurations using the twenty naturally occurring amino acids for a polymer five amino acids long is over three million. Typical proteins are more than one-hundred amino acids in length.

In typical applications, a complex solution containing one or more substances to be characterized contacts a polymer array comprising polypeptides. The polypeptides of the

20   invention can be prepared by classical methods known in the art, for example, by using standard solid phase techniques. The standard methods include exclusive solid phase synthesis, partial solid phase synthesis methods, fragment condensation, classical solution synthesis and recombinant DNA technology (see Merrifield, (1963) Am. Chem. Soc. 85, 2149-2152). On solid phase, the synthesis is typically commenced from the C-terminal end of

25   the peptide using an alpha-amino protected resin. A suitable starting material can be prepared, for instance, by attaching the required alpha-amino acid to a chloromethylated resin, a hydroxy-methyl resin or a benzhydrylamine resin.

The alpha-amino protecting groups are those known to be useful in the art of stepwise

- 30 -

31

synthesis of peptides. Included are acyl type protecting groups, aromatic urethane type protecting groups, aliphatic urethane protecting groups and alkyl type protecting groups. The side chain protecting group remains intact during coupling and is not split off during the deprotection of the amino-terminus protecting group or during coupling. The side chain

5 protecting group must be removable upon the completion of the synthesis of the final peptide and under reaction conditions that will not alter the target peptide.

After removal of the alpha-amino protecting group, the remaining protected amino acids are coupled stepwise in the desired order. An excess of each protected amino acid is generally used with an appropriate carboxyl group activator such as

10 dicyclohexylcarbodiimide (DCC) in solution, for example, in methylene chloride, dimethyl formamide (DMF) mixtures.

These procedures can also be used to synthesize peptides in which amino acids other than the twenty naturally occurring, genetically encoded amino acids are substituted at one, two, or more positions of any of the compounds of the invention. For instance,

15 naphthylalanine can be substituted for tryptophan, facilitating synthesis. Other synthetic amino acids that can be substituted into the peptides of the present invention include L-hydroxypropyl, L-3, 4-dihydroxyphenylalanyl, d-amino acids such as L-d-hydroxylysyl and D-d-methylalanyl, L-α-methylalanyl and β-amino acids non-naturally occurring synthetic amino acids can also be incorporated into the peptides of the present invention (see Roberts *et*

20 *al.*, (1983) Peptide Synthesis 5, 341-449).

One can replace the naturally occurring side chains of the twenty genetically encoded amino acids (or D amino acids) with other side chains, for instance with groups such as alkyl, lower alkyl, cyclic four, five, six, to seven-membered alkyl, amide, amide lower alkyl, amide di(lower alkyl), lower alkoxy, hydroxy, carboxy and the lower ester derivatives thereof, and

25 with four, five, six, to seven-membered heterocyclic. In particular, proline analogs in which the ring size of the proline residue is changed from five members to four, six or seven members can be employed. Cyclic groups can be saturated or unsaturated, and if unsaturated, can be aromatic or non-aromatic. Heterocyclic groups preferably contain one or more

- 31 -

32

nitrogen, oxygen, and/or sulphur heteroatoms. Examples of such groups include the furazanyl, furyl, imidazolidinyl, imidazolyl, imidazolinyl, isothiazolyl, isoxazolyl, morpholinyl, oxazolyl, piperazinyl, piperidyl, pyranyl, pyrazinyl, pyrazolidinyl, pyrazolinyl, pyrazolyl, pyridazinyl, pyridyl, pyrimidinyl, pyrrolidinyl, pyrrolinyl, pyrrolyl, thiadiazolyl,

5   thiazolyl, thienyl, thiomorpholinyl and triazolyl. These heterocyclic groups can be substituted or unsubstituted. Where a group is substituted, the substituent can be alkyl, alkoxy, halogen, oxygen, or substituted or unsubstituted phenyl.

One can also readily modify the peptides of the instant invention by phosphorylation (see Bannwarth *et al.*, (1996) Biorg. Med. Chem. Let. 6, 2141-2146) and other methods for

10   making peptide derivatives of the compounds of the present invention are described in Hruby *et al.*, (1990) Biochem. J. 268, 249-262). Thus, the peptide compounds of the invention also serve as a basis to prepare peptide mimetics with similar biological activity. The array can also comprise peptide mimetics with the same or similar desired biological activity as the corresponding peptide compound but with more favorable activity than the peptide with

15   respect to solubility, stability, and susceptibility to hydrolysis and proteolysis (see Morgan *et al.*, (1989) Ann. Rep. Med. Chem. 24, 243-252).

Peptides suitable for use in this embodiment generally include those peptides, for example, ligands, that bind to a receptor, such as seven transmembrane proteins. Such peptides typically comprise about 150 amino acid residues or less and, more preferably, about

20   100 amino acid residues or less.

The peptides of the present invention may exist in a cyclized form with an intramolecular disulfide bond between the thiol groups of the cysteines. Alternatively, an intermolecular disulfide bond between the thiol groups of the cysteines can be produced to yield a dimeric (or higher oligomeric) compound. One or more of the cysteine residues may

25   also be substituted with a homocysteine. Other embodiments of this invention provide for analogs of these disulfide derivatives in which one of the sulfurs has been replaced by a CH_2 group or other isostere for sulfur. These analogs can be made via an intramolecular or intermolecular displacement, using methods known in the art.

- 32 -

33

H. Methods to Identify Agents that Modulate Expression of DORs.

Another embodiment of the present invention provides methods for identifying agents that modulate the expression of a nucleic acid encoding any one of the DOR proteins of the invention such as any protein having the amino acid sequence depicted in SEQ ID NO: 2, 4, 6,

5   8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98. Such assays may utilize any available means of monitoring for changes in the expression level of the nucleic acids of the invention. As used herein, an agent is said to modulate the expression of a nucleic acid of the invention, for instance a nucleic acid encoding any one of the proteins

10   having the amino acid sequence depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98, if it is capable of up- or down-regulating expression of the nucleic acid in a cell.

In one assay format, cell lines that contain reporter gene fusions between the open

15   reading frame of any one of the nucleotides depicted in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 and 97 and any assay fusion partner may be prepared. Numerous assay fusion partners are known and readily available including the firefly luciferase gene and the gene encoding chloramphenicol acetyltransferase (Alam *et*

20   *al.,* (1990) Anal. Biochem. 188, 245-254). Cell lines containing the reporter gene fusions are then exposed to the agent to be tested under appropriate conditions and time. Differential expression of the reporter gene between samples exposed to the agent and control samples identifies agents which modulate the expression of a nucleic acid encoding at least one of the proteins having the sequence depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24,

25   26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98.

Additional assay formats may be used to monitor the ability of the agent to modulate the expression of a nucleic acid encoding at least one protein of the invention selected from

- 33 -

34

the group of proteins having SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98. For instance, mRNA expression may be monitored directly by hybridization to the nucleic acids of the invention. Cell lines are exposed to the

5   agent to be tested under appropriate conditions and time and total RNA or mRNA is isolated by standard procedures such those disclosed in Sambrook *et al.*, (1985) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory Press.

Probes to detect differences in RNA expression levels between cells exposed to the agent and control cells may be prepared from the nucleic acids of the invention. It is

10   preferable, but not necessary, to design probes which hybridize only with target nucleic acids under conditions of high stringency. Only highly complementary nucleic acid hybrids form under conditions of high stringency. Accordingly, the stringency of the assay conditions determines the amount of complementary nucleotides which should exist between two nucleic acid strands in order to form a hybrid. Stringency should be chosen to maximize the

15   difference in stability between the probe:target hybrid and potential probe:non-target hybrids.

Probes may be designed from the nucleic acids of the invention through methods known in the art. For instance, the G+C content of the probe and the probe length can affect probe binding to its target sequence. Methods to optimize probe specificity are commonly available in Sambrook *et al.*, (1985) Molecular Cloning - A Laboratory Manual, Cold Spring

20   Harbor Laboratory Press; or Ausubel *et al.*, (1995) Current Protocols in Molecular Biology, Greene Publishing Company.

Hybridization conditions are modified using known methods, such as those described by Sambrook *et al.*, (1985) and Ausubel *et al.*, (1995) as required for each probe. Hybridization of total cellular RNA or RNA enriched for polyA+ RNA can be accomplished

25   in any available format. For instance, total cellular RNA or RNA enriched for polyA RNA can be affixed to a solid support and the solid support exposed to at least one probe comprising at least one, or part of one of the sequences of the invention under conditions in which the probe will specifically hybridize. Alternatively, nucleic acid fragments comprising

- 34 -

at least one, or part of one of the sequences of the invention can be affixed to a solid support, such as a porous glass wafer. The glass wafer can then be exposed to total cellular RNA or polyA RNA from a sample under conditions in which the affixed sequences will specifically hybridize. Such glass wafers and hybridization methods are widely available, for example, those disclosed by Beattie (WO 95/11755). By examining for the ability of a given probe to specifically hybridize to an RNA sample from an untreated cell population and from a cell population exposed to the agent, agents which up- or down-regulate the expression of a nucleic acid encoding at least one protein having the amino acid sequence depicted in SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98 are identified.

Hybridization for qualitative and quantitative analysis of mRNA may also be carried out by using a RNase Protection Assay (*i.e.*, RPA, see Ma *et al.*, (1996) Methods 10, 273-238). Briefly, an expression vehicle comprising cDNA encoding the gene product and a phage specific DNA dependent RNA polymerase promoter (*e.g.*, T7, T3 or SP6 RNA polymerase) is linearized at the 3' end of the cDNA molecule, downstream from the phage promoter, wherein such a linearized molecule is subsequently used as a template for synthesis of a labeled antisense transcript of the cDNA by *in vitro* transcription. The labeled transcript is then hybridized to a mixture of isolated RNA (*i.e.*, total or fractionated mRNA) by incubation at 45°C overnight in a buffer comprising 80% formamide, 40 mM Pipes, pH 6.4, 0.4 M NaCl and 1 mM EDTA. The resulting hybrids are then digested in a buffer comprising 40 µg/ml ribonuclease A and 2 µg/ml ribonuclease. After deactivation and extraction of extraneous proteins, the samples are loaded onto urea-polyacrylamide gels for analysis.

In another assay format, agents which effect the expression of the instant gene products, cells or cell lines would first be identified which express said gene products physiologically. Cells and cell lines so identified would be expected to comprise the necessary cellular machinery such that the fidelity of modulation of the transcriptional apparatus is maintained with regard to exogenous contact of agent with appropriate surface

transduction mechanisms and the cytosolic cascades. Further, such cells or cell lines would be transduced or transfected with an expression vehicle (*e.g.*, a plasmid or viral vector) construct comprising an operable non-translated 5'-promoter containing end of the structural gene encoding the instant gene products fused to one or more antigenic fragments, which are

5   peculiar to the instant gene products, wherein said fragments are under the transcriptional control of said promoter and are expressed as polypeptides whose molecular weight can be distinguished from the naturally occurring polypeptides or may further comprise an immunologically distinct tag. Such a process is well known in the art (see Maniatis *et al.*, (1982) Molecular Cloning - A Laboratory Manual, Cold Spring Harbor Laboratory Press).

10   Cells or cell lines transduced or transfected as outlined above would then be contacted with agents under appropriate conditions; for example, the agent comprises an acceptable excipient and is contacted with cells comprised in an aqueous physiological buffer such as phosphate buffered saline (PBS) at physiological pH, Eagles balanced salt solution (BSS) at physiological pH, PBS or BSS comprising serum or conditioned media comprising PBS or

15   BSS and/or serum incubated at 37°C. Said conditions may be modulated as deemed necessary by one of skill in the art. Subsequent to contacting the cells with the agent, said cells will be disrupted and the polypeptides from disrupted cells are fractionated such that a polypeptide fraction is pooled and contacted with an antibody to be further processed by immunological assay (*e.g.*, ELISA, immunoprecipitation or Western blot). The pool of

20   proteins isolated from the "agent contacted" sample will be compared with a control sample where only the excipient is contacted with the cells and an increase or decrease in the immunologically generated signal from the "agent contacted" sample compared to the control will be used to distinguish the effectiveness of the agent.

25   **I. Methods to Identify Agents that Modulate Activity of DORs**

Another embodiment of the present invention provides methods for identifying agents that modulate at least one activity of a protein of the invention such as any one of the proteins having the amino acid sequence of SEQ ID NO: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26,

37

28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96 and 98. Such methods or assays may utilize any means of monitoring or detecting the desired activity.

In one format, the relative amounts of a protein of the invention between a cell population that has been exposed to the agent to be tested compared to an un-exposed control cell population may be assayed. In this format, probes such as specific antibodies are used to monitor the differential expression of the protein in the different cell populations. Cell lines or populations are exposed to the agent to be tested under appropriate conditions and time. Cellular lysates may be prepared from the exposed cell line or population and a control, unexposed cell line or population. The cellular lysates are then analyzed with the probe.

Antibody probes are prepared by immunizing suitable mammalian hosts in appropriate immunization protocols using the peptides, polypeptides or proteins of the invention if they are of sufficient length, or if desired, required to enhance immunogenicity, conjugated to suitable carriers. Methods for preparing immunogenic conjugates with carriers such as BSA, KLH, or other carrier proteins are well known in the art. In some circumstances, direct conjugation using, for example, carbodiimide reagents may be effective; in other instances linking reagents such as those supplied by Pierce Chemical Co., may be desirable to provide accessibility to the hapten. The hapten peptides can be extended at either the amino or carboxy terminus with a cysteine residue or interspersed with cysteine residues, for example, to facilitate linking to a carrier. Administration of the immunogens is conducted generally by injection over a suitable time period and with use of suitable adjuvants, as is generally understood in the art. During the immunization schedule, titers of antibodies are taken to determine adequacy of antibody formation.

While the polyclonal antisera produced in this way may be satisfactory for some applications, for some applications, use of monoclonal preparations is preferred. Immortalized cell lines which secrete the desired monoclonal antibodies may be prepared using the standard method of Kohler & Milstein, (1975) Nature 256, 495-497 or modifications which effect immortalization of lymphocytes or spleen cells, as is generally known. The

- 37 -

immortalized cell lines secreting the desired antibodies are screened by immunoassay in which the antigen is the peptide hapten, polypeptide or protein. When the appropriate immortalized cell culture secreting the desired antibody is identified, the cells can be cultured either *in vitro* or by production in ascites fluid.

5   The desired monoclonal antibodies are then recovered from the culture supernatant or from the ascites supernatant. Fragments of the monoclonal or polyclonal antisera which contain the immunologically significant portion can be used as antagonists, as well as the intact antibodies. Use of immunologically reactive fragments, such as the Fab, Fab' of $F(ab')_2$ fragments is often preferable, as these fragments are generally less immunogenic than the
10   whole immunoglobulin.

The antibodies or fragments may also be produced, using current technology, by recombinant means. Antibody regions that bind specifically to the desired regions of the protein can also be produced in the context of chimeras with multiple species origin, particularly humanized antibodies.

15   Agents that are assayed in the above method can be randomly selected or rationally selected or designed. As used herein, an agent is said to be randomly selected when the agent is chosen randomly without considering the specific sequences involved in the association of the a protein of the invention alone or with its associated substrates, binding partners, etc. An example of randomly selected agents is the use a chemical library or a peptide combinatorial
20   library, or a growth broth of an organism.

As used herein, an agent is said to be rationally selected or designed when the agent is chosen on a non-random basis which takes into account the sequence of the target site and its conformation in connection with the agent's action. Agents can be rationally selected or rationally designed by utilizing the peptide sequences to identify proposed binding motifs,
25   glycosylation and phosphorylation sites on the protein.

The agents of the present invention can be, as examples, peptides, small molecules, vitamin derivatives, as well as carbohydrates. A skilled artisan can readily recognize that there is no limit as to the structural nature of the agents of the present invention. Dominant-

- 38 -

negative proteins, DNA encoding these proteins, antibodies to these proteins, peptide fragments of these proteins or mimics of these proteins may be contacted with cells to affect function. "Mimic" as used herein refers to the modification of a region or several regions of a peptide molecule to provide a structure chemically different from the parent peptide but

5   topographically and functionally similar to the parent peptide (see Meyers, (1995) Molecular Biology & Biotechnology, VCH Publishers).

The peptide agents of the invention can be prepared using standard solid phase (or solution phase) peptide synthesis methods, as is known in the art. In addition, the DNA encoding these peptides may be synthesized using commercially available oligonucleotide

10   synthesis instrumentation and produced recombinantly using standard recombinant production systems. The production using solid phase peptide synthesis is necessitated if non-gene-encoded amino acids are to be included.

Another class of agents of the present invention are antibodies immunoreactive with critical positions of proteins of the invention. Antibody agents are obtained by immunization

15   of suitable mammalian subjects with peptides, containing as antigenic regions, those portions of the protein intended to be targeted by the antibodies.

J. Transgenic Organisms

Transgenic insects containing mutant, knock-out or modified genes corresponding to

20   any one of the cDNA sequences depicted in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 and 97 are also included in the invention. Transgenic insects are genetically modified insects into which recombinant, exogenous or cloned genetic material has been experimentally transferred. Such genetic material is often

25   referred to as a "transgene". The nucleic acid sequence of the transgene, in this case a form of any one of the sequences depicted in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95 and 97, may be integrated either at a locus of a genome

- 39 -

where that particular nucleic acid sequence is not otherwise normally found or at the normal locus for the transgene. The transgene may consist of nucleic acid sequences derived from the genome of the same species or of a different species than the species of the target insect.

The term "germ cell line transgenic insect" refers to a transgenic insect in which the genetic alteration or genetic information was introduced into a germ line cell, thereby conferring the ability of the transgenic insect to transfer the genetic information to offspring. If such offspring in fact possess some or all of that alteration or genetic information, then they too are transgenic insects.

The alteration or genetic information may be foreign to the species of insect to which the recipient belongs, foreign only to the particular individual recipient, or may be genetic information already possessed by the recipient. In the last case, the altered or introduced gene may be expressed (*i.e.*, over-expression and knock-out) differently than the native gene.

Transgenic insects can be produced by a variety of different methods including P element-mediated transformation by microinjection (see, *e.g.*, Rubin & Spradling, (1982) Science 218, 348-353; Orr & Sohal, (1993) Arch. Biochem. Biophys. 301, 34-40), transformation by microinjection followed by transgene mobilization (Mockett *et al.*, (1999) Arch. Biochem. Biophys. 371, 260-269), electroporation (Huynh & Zieler, (1999) J. Mol. Biol. 288, 13-20) and through the use of baculovirus (Yamao *et al.*, (1999) Genes Dev. 13, 511-516. Furthermore, the use of adenoviral vectors to direct expression of a foreign gene to olfactory neuronal cells can also be used to generate transgenic insects (see, *e.g.*, Holtmaat *et al.*, (1996) Brain. Res. Mol. Brain Res. 41, 148-156).

A number of recombinant or transgenic insects have been produced, including those which over-express superoxide dismutase (Mockett *et al.*, (1999) Arch. Biochem. Biophys. 371, 260-269); express Syrian hamster prion protein (Raeber *et al.*, (1995) Mech. Dev. 51, 317-327); express cell-cycle inhibitory peptide aptamers (Kolonin & Finley (1998) Proc. Natl. Acad. Sci. USA 95, 14266-14271); and those which lack expression of the putative ribosomal protein S3A gene (Reynaud *et al.*, (1997) Mol. Gen. Genet. 256, 462-467).

While insects remain the preferred choice for most transgenic experimentation, in

some instances it is preferable or even necessary to use alternative animal species. Transgenic procedures have been successfully utilized in a variety of animals, including mice, rats, sheep, goats, pigs, dogs, cats, monkeys, chimpanzees, hamsters, rabbits, cows and guinea pigs (see, *e.g.*, Kim *et al.*, (1997) Mol. Reprod. Dev. 46, 515-526; Houdebine, (1995)

5   Reprod. Nutr. Dev. 35, 609-617; Petters, (1994) Reprod. Fertil. Dev. 6, 643-645; Schnieke *et al.*, (1997) Science 278, 2130-2133; and Amoah, (1997) J. Anim. Sci. 75, 578-585).

The method of introduction of nucleic acid fragments into insect cells can be by any method which favors co-transformation of multiple nucleic acid molecules. For instance, *Drosophila* embryonic Schneider line 2 (S2) cells can be stably transfected as previously

10   described (Schneider, (1972) J. Embryol. Exp. Morphol. 27, 353-365). Detailed procedures for producing transgenic insects are readily available to one skilled in the art (see Rubin & Spradling, (1982) Science 218, 348-353; Orr & Sohal, (1993) Arch. Biochem. Biophys. 301, 34-40, herein incorporated by reference in their entirety).

15   **K. Uses for Agents that Modulate at Least One Activity of DORs**

1. Introduction.

Organisms, including insects, are continually exposed to a great number of volatiles released by other organisms as well as by other aspects of their environment. The olfactory receptor genes of the present invention play an important role in the detection and processing

20   of these chemical stimuli, some of which have been implicated in initiating and modulating host-seeking and other behaviors, such as mating behaviors (see, for example, Roth, (1951) Ann. Entomol. Soc. Am. 44, 59-74; Jones *et al.*, (1976) Ent. Exp. Appn. 19, 19-22; Gillies, (1980) Bull. Ent. Res. 70, 525-532; Kline *et al.*, (1991) J. Med. Entomol. 28, 254-258). For a recent, thorough review of the many practical applications of the present invention (see Karg

25   & Suckling, (1999) Applied aspects of insect olfaction, in: Hansson (ed.), Insect Olfaction, Springer, which is incorporated by reference in its entirety).

Most importantly, the DOR genes of the present invention may be used to track down odor receptor genes in insects that damage crops or transmit diseases. The present invention

- 41 -

provides the tools and methodologies for finding specific compounds that interfere with the insects' ability to detect odors.

Of course, the present invention has important implications for improved methods of using pheromones and other semiochemicals for pest control. In addition, recent

5 advancements in many other fields have greatly increased the variety of additional technologies for which the present invention also has significant applications. Examples of such'advancements include, but are not limited to the following: i) the development and application of new techniques of chemical identification and synthesis; ii) new chemical release techniques; iii) more sophisticated application technologies; and iv) more detailed

10 information about the behavior of specific organisms.

While not wishing to be bound by the specific embodiments discussed herein, the following sections provide an overview of the wide variety of applications for which the present invention may be employed.

2. Definitions.

15 As used herein, the term "allomones" refers to any chemical substance produced or acquired by an organism that, when it contacts an individual of another species, evokes in the receiver a behavioral or developmental reaction adaptively favorable to the transmitter.

As used herein, the term "host" refers to any organism on which another organism depends for some life function. Examples of hosts include, but are not limited to, humans

20 which may serve as a host for the feeding of certain species of mosquito and the leaves of soybeans (*Glycine max*(L.)) which may act as hosts for the oviposit of the green cloverworm (*Plathypena scabra* (F.)).

As used herein, the term "kairomones" refers to any of a heterogeneous group of chemical messengers that are emitted by organisms of one species but benefit members of

25 another species. Examples include, but are not limited to, attractants, phagostimulants, and other substances that mediate the positive responses of, for example, predators to their prey, herbivores to their food plants, and parasites to their hosts. Kairomones suitable for the purposes of the invention and methods of obtaining them are described, for example, Science

43

(1966) 154, 1392-93; Hedin, (1985) Bioregulators for Pest Control, American Chemical Society, Washington, 353-366.

As used herein, the term "pheromone" refers to a substance, or characteristic mixture of substances, that is secreted and released by an organism and detected by a second organism

5 of the same or a closely related species, in which it causes a specific reaction, such as a definite behavioral reaction or a developmental process. Examples include, but are not limited to, the mating pheromones of fungi and insects. More than a thousand moth sex pheromones (Toth *et al.*, (1992) J. Chem. Ecol. 18, 13-25 ; Arn *et al.*, (1998) Appl. Entomol. Zoo. 33, 507-511) and hundreds of other pheromones have now been identified, including aggregation

10 pheromones from beetles and other groups of insects. Various compositions, including resins and composite polymer dispensers, have been developed for the controlled release of pheromones have been developed (see, *e.g.*, U.S. Patent No. 5,750,129 & 5,504,142).

As used herein, the term "semiochemical" refers to any chemical substance that delivers a message or signal from one organism to another. Examples of such chemicals

15 include, but are not limited to, pheromones, kairomones, oviposition deterrents, or stimulants, and a wide range of other classes of chemicals (see, for example, Nordlund, (1981) Semiochemicals: A review of the terminology, in: Nordlund *et al.*, (ed.) Semiochemicals: Their Role in Pest Control, John Wiley; Howse *et al.*, (1998) Insect Pheromones and Their Use in Pest Management, Chapman & Hall, London).

20 As used herein, the term "synomones" refers to any chemical substance which benefits both the emitter and receiver. Examples include, but are not limited to, compounds involved in floral attraction of pollinators and species-isolating mechanisms, such as sex pheromones of related species, where an inhibitor often functions to prevent mating among sympatric species.

As used herein, the term "volatile" refers to a chemical which evaporates readily at

25 those temperatures and pressures which are considered the relevant temperatures and pressures for the reference organism of interest.

3. As Tools for Further Scientific Research.

<u>Identification of Olfactory Receptor Genes in Other Organisms.</u> The algorithms of the

ЧЧ

present invention may be used directly to search for olfactory receptor genes in other organisms, as explained elsewhere herein.

Alternatively, nucleic acid probes or primers may be designed based on the DOR genes of the present invention. Such probes or primers may be used to identify and isolate 5 olfactory receptor genes in other organisms. Methods of creating and using the necessary nucleic acid probes and primers are discussed elsewhere herein.

The highest probability of success in locating olfactory genes in other organisms using the DOR genes of the present invention will most likely occur by using a boot-strapping or leap-frogging method. Such methods involve first probing organisms most related to fruit 10 flies and successively progressing to more unrelated organisms, using the most newly identified olfactory receptor genes to identify similar genes in the next, more unrelated, insect of interest. Thus, the first organisms to probe with the DOR genes of the present invention most preferably may be other flies from the order *Diptera* (*i.e.*, the two-winged or true flies). Examples of suitable flies include, but are not limited to, the tsetse fly, horse fly, house fly, 15 bluebottle fly, hover fly and mosquito. *Dipterans* which transmit diseases causing serious health problems are of particular interest (*e.g.*, horse fly, tsetse fly, mosquito).

After the identification of olfactory receptor genes in various *Diptera* insects, the next organisms to probe most preferably may be from orders within the same subclass as *Diptera*. Finally, the next insects to use would be those from orders not within the same subclass as 20 *Diptera*.

The insects which cause substantial health risks, crop damage, or other significant damage (*e.g.*, to housing structure or cotton clothing) may be the most desirable targets for such studies. Examples of such insects include, but are not limited to, green cloverworm, Mexican bean beetle, potato leafhopper, corn earworm, green stink bug, northern corn 25 rootworm, western corn rootworm, cutworms, wireworms, thrips, fleas, aphids (*e.g.*, pea aphid, spotted alfalfa aphid), European corn borer, fall armyworm, southwestern corn borer, grasshoppers, Japanese beetle, termites, leafhoppers (*e.g.*, potato leafhopper, three-cornered alfalfa hopper), stink bugs, crickets, Hessian fly, greenbugs and weevils (*e.g.*, alfalfa weevil,

- 44 -

bollweevil).

Olfactory receptor genes identified by this process may then be used to screen non-Insecta organisms for olfactory receptor genes. Organisms of interest may include, but be limited to, mites, ticks, spiders, nematodes, centipedes, mice, rats, salmon, pigeons, dogs,

5 horses and humans.

Genetic Manipulations. The tools and methodologies of the present invention may be used by neurobiologists to probe more complex workings of an organism's response system, including those of a mammal's brain.

Knock-outs. By systematically knocking out the olfactory receptor genes of the

10 present invention and observing the effects on odor sensitivity and behavior, researchers will be able to piece together a wiring diagram of the olfactory system of the fruit fly.

The term "knock-out" generally refers to mutant organisms which contain a null allele of a specific gene. Methods of making knock-out or disruption transgenic animals, especially mice, are generally known by those skilled in the art and are discussed herein and elsewhere

15 (see, for example, the section herein entitled Transgenic Organisms and the following: Manipulating the Mouse Embryo, (1986) Cold Spring Harbor Laboratory Press; Capecchi, (1989) Science 244, 1288-1292; Li et al., (1995) Cell 80, 401-411; U.S. Patent No. 5,981,830 & 5,789,654, each of which is incorporated herein by reference.

Parallel studies may be conducted in other organisms by using the olfactory receptor

20 genes and the methods of the present invention to identify the olfactory receptor genes of other organisms and then creating knock-outs for the olfactory receptor genes of those organisms.

Disabling Genes. Using the olfactory receptor genes of the present invention, it is now possible to selectively disable specific DOR genes and look for changes in odor response and

25 behavior. Parallel studies may be conducted in other organisms by using the olfactory receptor genes and the methods of the present invention to identify the olfactory receptor genes of other organisms and then disabling olfactory receptor genes of those organisms.

Methods of disabling genes are generally known by those skilled in the art. An

example of an effective disabling modification would be a single nucleotide deletion occurring at the beginning of a olfactory receptor gene that would produce a translational reading frameshift. Such a frameshift would disable the gene, resulting in non-expressible gene product and thereby disrupting functional protein production by that gene. Protease production by the gene could be disrupted if the regulatory regions or the coding regions of the protease genes are disrupted.

In addition to disabling genes by deleting nucleotides, causing a transitional reading frameshift, disabling modifications would also be possible by other techniques including insertions, substitutions, inversions or transversions of nucleotides within the gene's DNA that would effectively prevent the formation of the protein coded for by the DNA.

It is also within the capabilities of one skilled in the art to disable genes by the use of less specific methods. Examples of less specific methods would be the use of chemical mutagens such as hydroxylamine or nitrosoguanidine or the use of radiation mutagens such as gamma radiation or ultraviolet radiation to randomly mutate genes, such as the DOR genes of the present invention. Such mutated strains could, by chance, contain disabled olfactory receptor genes such that the genes are no longer capable of producing functional proteins for any one or more of the domains. The presence of the desired disabled genes could be detected by routine screening techniques. For further guidance, see U.S. Patent No. 5,759,538.

Over-expression. Using the olfactory receptor genes of the present invention, it is now possible to selectively over-express specific DOR genes and look for changes in odor response and behavior. Parallel studies may be conducted in other organisms by using the olfactory receptor genes and the methods of the present invention to identify the olfactory receptor genes of other organisms and then overexpress the olfactory receptor genes of those organisms.

Methods of overexpressing genes are generally known by those skilled in the art. For examples of producing cells which overexpress specific genes, see, for example, U.S. Patent Numbers 5,905,146; 5,849,999; 5,859,311; 5,602,309; 5,952,169 and 5,772,997 (HER2 receptor).

Modulating or Inhibiting Expression. Using the olfactory receptor genes of the present invention, it is now possible to selectively modulate or inhibit specific DOR genes using antisense oligomers which specifically hybridize with the DNA or RNA encoding the DOR genes. One skilled in the art could so modulate or inhibit the expression of the DOR genes and detect for changes in odor response and behavior. Parallel studies may be conducted in other organisms by using the olfactory receptor genes and the methods of the present invention to identify the olfactory receptor genes in other organisms and then use antisense oligers to the olfactory receptor genes of those organisms. Methods for inhibiting expression of genes, especially genes coding for receptor genes, using antisense constructs, including generation of antisense sequences *in situ* are described, for example, in U.S. Patent Numbers 5,856,099; 5,556,956; 5,716,846; 5,135,917 and 6,004,814.

Other methods that can be used to inhibit expression of an endogenous gene are applicable to the present invention. For example, formation of a triple helix at an essential region of a duplex gene serves this purpose. The triplex code, permitting design of the proper single stranded participant is also known in the art. (See H. E. Moser, *et al.,* (1987) Science 238: 645-650 and M. Cooney, *et al.,* (1988) Science 241: 456-459). Regions in the control sequences containing stretches of purine bases are particularly attractive targets. Triple helix formation along with photocrosslinking is described, *e.g.*, in Praseuth *et al.,* (1988) Proc. Natl Acad. Sci. USA 85:1349-1353.

Studying Behavior. The present invention is useful for studying the developmental aspects of the olfactory receptor genes which appear to be active at different times during development. Such studies may help organize the olfactory systems in various organisms and may help explain the behavior of various organisms.

The tools and methodologies of the present invention may be used to study the influence of environmental conditions on pheromone communication. For example, newly identified olfactory receptor genes may be used to study the effects of different rearing temperatures and light regimes (selected to mimic those occurring in the spring and summer growing seasons) on the response of various *Lepidoptera* insects, such as the cabbage looper

- 47 -

moth (*Trichoplusia ni* (Hubner)). For a description of the methods which might be used for such a study, see, for example, Grant *et al.*, (1996) Physiol. Entomol. 21, 59-63.

4. For Organism Detection, Monitoring and Control.

General Pest Management. The olfactory receptor genes identified herein and identified using the methods of the present invention may be used to identify compounds which may be used for pest management. It is especially desirable to utilize various aspects of the present invention for pest management related to crop protection.

The application of pheromones is now firmly established as a key component of pest management and control, especially within the framework of integrated pest management (IPM). An object of organism control is to modulate an organism's behavior or activity so as to reduce the irritation, sickness, or death of the host (*e.g.*, a plant host), or to decrease the general health and proliferation of the organism.

For example, the propagation of a mouse population in a given area of actual or potential mice infestation may be prevented or inhibited by treating such an area with an effective amount of male mouse pheromones, wherein such pheromones have male mouse aversion signaling properties (see, *e.g.*, U.S. Patent No. 5,252,326).

Insect Repellents and Insecticides. The present invention provides the tools and methodologies useful for identifying compounds which modulate insect behavior by exploiting the sensory capabilities of the target insect. For example, attempts have been made to describe and synthesize the complex interactions which underlie host-seeking behavior in mosquitoes. Using the methods and olfactory receptor genes of the present invention, it is possible to design specific compounds which target mosquito olfactory receptor genes. Thus, the present invention provides the ability to alter or to eliminate the orientation and feeding behaviors of mosquitoes and thereby have a positive impact on world health by controlling mosquito-borne diseases, such as malaria.

Mosquito olfactory receptor genes may be identified and/or targeted using various aspects of the present invention. For example, the olfactory receptor genes of the present invention may be used to design probes as discussed elsewhere herein for the identification

and characterization of mosquito olfactory receptor genes. Alternatively, the algorithm of the present invention may be used to identify mosquito olfactory receptor genes in the genetic databases for mosquitoes. Once the mosquito olfactory receptor genes are identified, then various screening methods described elsewhere herein, such as the high throughput assays

5   discussed elsewhere herein, may be used to identify synthetic and natural compounds which may modulate the behavior of the insect.

Mating Enhancement and Disruption. The olfactory receptor genes identified herein and identified using the methods of the present invention may be used to identify compounds which interfere with the orientation and mating of a wide range of organisms, including

10   insects. Thus, the present invention enables the identification of compositions which disrupt insect mating by selective inhibition of specific receptor genes involved in mating attraction (see, e.g., U.S. Patent No. 5,064,820).

Animal Repellants. The olfactory receptor genes identified herein and identified using the methods of the present invention may be used to identify compounds which may be used

15   as animal repellants. Such compositions may be used to repel both predatory and non-predatory animals (see, e.g., U.S. Patent No. 4,668,455).

6. Organism Attraction.

Insect Attractants. The olfactory receptor genes identified herein and identified using the methods of the present invention may be used to identify compounds which attract

20   specific insects to a particular location (see, e.g., U.S. Patent No. 4,880,624 & 4,851,218).

For example, aspects of the present invention may to used in various methods which reduce or eliminate the levels of particular insect pests, such as mosquitoes and tsetse flies. As a particular example, insect traps can be created wherein the pheromone attracts a particular insect, like the tsetse fly, and the insect so attracted dies in the trap. In this way, the

25   population of tsetse flies may be reduced or eliminated in a particular area.

The insect attractant compositions so identified may also be combined with an insecticide, for example as an insect bait in microencapsulated form. Alternatively, or in addition, the insect attractant composition may be placed inside an insect trap, or in the

- 49 -

vicinity of the entrance to an insect trap.

In addition to killing insects, the trapping of insects is often very important for estimating or calculating how many insects of a particular type are feeding within a specific area. Such estimates are used to determine where and when insecticide spraying should be

5   commenced and terminated.

Insect traps which may be used are, for example, those as described in PCT/BG93/01442 and U.S. Patent No. 5,713,153. Specific examples of insect traps include, but are not limited to, the Gypsy Moth Delta Trap®, Boll Weevil Scout Trap®, Jackson trap, Japanese beetle trap, McPhail trap, Pherocon 1C trap, Pherocon II trap, Perocon AM trap and

10   Trogo trap.

Kairomones may be used as an attractancy for the enhancement of the pollination of selected plant species.

Attractant compositions which demonstrate biological activity toward one sex which is greater than toward the opposite sex may be useful in trapping one sex of a specific

15   organism over another. For example, a composition may be a highly effective attractant for male apple ermine moths (*Yponomeuta malinellus* (Zeller)) and not so effective an attractant for female apple ermine moths. By attracting adult males to field traps, the composition provides a means for detecting, monitoring, and controlling this agricultural pest (see, *e.g.*, U.S. Patent No. 5,380,524).

20   <u>Attracting Predators and Parasitoids</u>. The olfactory receptor genes of the present invention and the olfactory receptor genes identified using the methods of the present invention may also be used to identify chemicals which attract various predators and parasitoids. Attracting the predators and parasitoids which attack certain pests offers an alternative method of pest management.

25   <u>Animal Attractants</u>. The olfactory receptor genes identified herein and those identified by the methods of the present invention may be used to identify chemicals which attract household domesticated animals. For example, a pheromone-containing litter preparation may attract the animals and absorb liquids and liquid-containing waste released by the

attracted animal (see, *e.g.*, U.S. Patent No. 5,415,131).

Synthetic Perfumes. A "perfume" or a "fragrance composition" is a specific pleasantly odorous cosmetic composition for topical application to an individual. The olfactory receptor genes identified herein and those identified by the methods of the present invention may be used to identify chemicals which may be produced and used as synthetic perfumes. Such perfumes may be used to disguise odors or enhance attraction between humans (see, *e.g.*, U.S. Patent No. 5,278,141).

7. Pharmaceuticals. The olfactory receptor genes identified herein and those identified using the methods of the present invention may be used to identify pharmaceutical compounds useful for altering the behavior and physiology of animals. Examples of such compounds include, but are not limited to, certain Androstene steroids that effectuate a change in human hypothalamic function (see, *e.g.*, U.S. Patent No. 5,969,168).

8. Industrial Applications. The olfactory receptor genes identified by the methods of the present invention may be used for a number of different industrial applications including, but not limited to the following:

a) Identification of appetite suppressant compounds and using same to suppress and/or control appetite.

b) Trapping odors of a specific type.

c) As Biosensors.

1) Explosive and drug detectors. The detectors may be synthetic, such as biologically-inspired robotic sensors, or biological sensors, such as sniffing dogs which are especially sensitive to certain odors.

2) Population of olfactory receptor genes expressed in cell culture. Olfactory receptor genes can be introduced into a cell line and the transformed cells maintained in culture through multiple generations. By creating specific cell lines which express multiple olfactory genes at once, it would be possible to use such cell cultures to investigate how odorants interact with odorant receptor genes. Thus, the present invention provides methods for identifying odorant fingerprints, wherein such methods include contacting a series of cells

5ʔ

containing and expressing known odor receptor genes with a desired sample, and determining the type and quantity of the odorant ligands present in the sample (see, *e.g.*, U.S. Patent No. 5,993,778). As discussed elsewhere herein, the interaction of substances with the receptors can be identified using appropriate labels, such as those provided by luciferase, the jellyfish

5 green fluorescent protein (GFP) or β-galactosidase.

3) Biochip Arrays. As discussed elsewhere herein, biochip arrays of odorant receptor genes can be generated. The arrays may be used to detect olfactory receptor ligands via an appropriate marker or via a chemical or electrical signal. Arrays may be designed for specific purposes, such as, but not limited to, detecting perfumes, explosives, drugs, pollutants, and

10 toxins.

d) Training organisms to conduct certain tasks. Examples include, but are not limited to, the following:

1) Training mice to pull guide line for stringing fiber optic cable through existing conduit holding copper wire.

15 2) Training mice to find their way through a maze based on smell (see, *e.g.,* Otto *et al.,* (1991) Hippocampus 1, 181-192; Granger *et al.,* (1991) Psych. Science 2, 116-118).

3) Improving the orientation and homing performance of pigeons (see, *e.g.,* Wiltschko, (1996) J. Exp. Biol. 199, 113-119) and fish (see, *e.g.,* Cao et al. (1998) Proc. Natl. Acad. Sci. USA 95(20):11987-11992).

20 4) Orient or reorient the behavior of worker bees of a rearing colony by incorporating a composition which includes one or more pheromones which elicits particular bee behavior towards the larvae. Thus, the beekeeper may orient or reorient the bees towards a particular activity such as, but not limited to, inducing improved acceptance of the larvae at the beginning of rearing, to increase the production of royal jelly, regulate the feeding of the

25 larvae as to favor the development of queen bees, etc. (see, *e.g.,* U.S. Patent No. 5,695,383).

Without further description, it is believed that one of ordinary skill in the art can, using the preceding description and the following illustrative examples, make and utilize the

compounds of the present invention and practice the claimed methods. The following

working examples therefore, specifically point out the preferred embodiments of the present

invention, and are not to be construed as limiting in any way the remainder of the disclosure.

5   **EXAMPLES**

Example 1: Identification of candidate olfactory receptor genes

In vertebrates and nematodes it is estimated that there are hundreds of olfactory

receptor genes, widely distributed in the genome (Buck & Axel, (1991) Cell 65, 175-187;

Troemel *et al.*, (1995) Cell 83, 207-218). With approximately 10% of the *Drosophila*

10   genome sequenced, it was likely that some of the *Drosophila* odorant receptor genes have

been sequenced. A two-step strategy was developed to identify odorant receptor genes from

the genomic database. First, a computer algorithm was designed to search the *Drosophila*

genomic sequence for open reading frames (ORFs) from candidate odorant receptor genes.

Second, RT-PCR was used to determine if transcripts from any of these ORFs were expressed

15   in olfactory organs. Finally, *in situ* hybridization was used to localize expression of DOR

genes.

Step 1: Computer algorithm for identification of GPCR genes. The algorithm used to

identify GPCR genes used statistical characterization of amino acid physico-chemical profiles

in combination with a non-parametric discriminant function. The key approach is to use the

20   information in the interplay between the local structure (transmembrane alpha helix) and the

global structure (repeated multiple domains) and characterize this information with concise

statistical variables. The algorithm was trained on a set of 100 putative GPCR sequences from

the GPCR database (GPCRDB) at http://swift.embl-heidelberg.de/7tm and a set of 100

random proteins selected from the SWISSPROT database (this training set was later

25   expanded, but that version was not used for the genes reported in this paper). In the first step,

three sets of descriptors were used to summarize the physico-chemical profiles of the

sequences. These were GES scale of hydropathy (Engelman *et al.*, (1986) Annu. Rev.

Biophys. Biophys. Chem. 15, 321-353), polarity (Brown, (1991) Molecular Biology Labfax,

- 53 -

54

Academic Press), and amino acid usage frequency. For the first two of these measurements, a sliding window profile was employed (White, (1994) Membrane Protein Structure, Oxford University Press) using a kernel of 15 amino acid constant function convoluted with a 16 amino acid Gaussian function. These profiles were then summarized with three statistics; the

5 periodicity (characterizing the quasi-periodic presence of the transmembrane domain), average derivative (characterizing the abrupt change between the transmembrane domain and non-transmembrane domain), and the variance of the derivative (also characterizing the abrupt change). GES periodicity, variance of polarity derivative, polarity periodicity and amino acid frequency were used as the four variables and each sequence was therefore characterized by

10 four variables. These four variables were used in a non-parametric linear discriminant function that was then optimized to separate the known GPCRs from random proteins in the training set. The same linear discriminant function with the scores derived from the training set was then used to screen the genomic database for candidate genes. The candidate sequences were given significance values by an odds ratio of the GPCRs and non-GPCRs computed using the

15 observed empirical distribution of the training set. More detailed information about the algorithm is available at http://www.neuron.org/cgi/content/full/22/2/327/dc1.

The computational screens used the genomic sequence data obtained by FTP from the Berkeley *Drosophila* Genome Project (BDGP, http://www.fruitfly.org, version 6/98). First, the ORFs of 300 bases or longer in all six frames were identified. Next, a program written to

20 identify GPCRs statistically by their physico-chemical profile was used to screen for candidate ORFs as described above. The number of possible candidates was reduced by comparing them to *Drosophila* codon usage tables (http://flybase.bio.indiana.edu, version 10). Candidate ORFs whose codon usage differed at a significance level of 0.0005 by the chi-square statistic were discarded from the candidate set. Using these screening steps, 34 candidate ORFs were

25 obtained.

Further analysis revealed that eight of the thirty-four candidate ORFs corresponded to genes of known function, for example a cyclic nucleotide-gated channel (Baumann *et al.*, (1994) EMBO J. 13, 5040-5050) and these ORFs were not further analyzed. Most of the

remaining ORFs encoded fewer than seven predicted transmembrane domains. The genomic DNA surrounding each of the computer-identified ORFs was therefore examined for the presence of neighboring ORFs encoding additional transmembrane domains to which the original ORFs might be spliced. *Drosophila* 5' and 3' intron-exon consensus splice sequences

5   were used in this analysis to help identify linked exons (Mount *et al.*, (1992) Nucleic Acids Res. 20, 4255-4262). This analysis yielded several genes that encoded seven-transmembrane-domain proteins (22A.1 and 22A.2).

Step 2: Sequence analysis of DOR olfactory genes. To determine if these two candidates were part of a larger family of genes encoding seven-transmembrane-domain

10   proteins, BLAST searches of the *Drosophila* genome database were conducted using the candidate gene sequences to identify related genes (Altschul *et al.*, (1990) J. Mol. Biol. 215, 403-410). The computer algorithms employed identified the ORFs for the second exons of 22A.1 and 22A.2, which encode transmembrane domains 1-4. These ORFs are on the BDGP P1 clone designated DS005342. The DS005342 sequence was examined around the initial

15   ORFs for neighboring ORFs which encoded additional potential transmembrane domains. Key to the identification of these neighboring ORFs was the presence of intron-exon consensus splice sequences: GTRAGT for the 5' end and HAG for the 3' end (Mount *et al.*, (1992) Nucleic Acids Res. 20, 4255-4262). 22A.1 and 22A.2 were found to have two other introns in corresponding locations, all of which had conserved splice sequences.

20   The amino acid sequences of 22A.1 and 22A.2 were used in searches of the *Drosophila* genome database using the tBLASTn program of the BDGP. These searches yielded partial sequences of other members of the DOR family. To complete the sequences of these genes, an analysis of the genomic DNA around each identified ORF was carried out as was done for 22A.1 and 22A.2, using the locations of conserved introns in the genes, the

25   intron consensus splice sequences, and the tBLASTn alignments as guides. Use of the genes identified in the second round as query sequences in tBLASTn searches and subsequent similar analysis of genomic DNA yielded the remaining genes. Additional searches of GenBank and SwissProt databases were performed with the NCBI (National Center for

56

Biotechnology Information) BLAST network.

The sequence alignment in Figure 3 is based on the alignments predicted by the tBLASTn program of the BDGP but was edited extensively. The 5' splice sequences for the most 3' introns of both 2F.1 and 47E.1 were unfavorable. It was assumed that these introns

5   were spliced nonetheless, as the resulting amino acid sequence displayed greater sequence identity to other DOR family members. If these introns were not spliced out, then the lengths of 2F.1 and 47E.1 would not be significantly altered from the lengths indicated in Figure 3. 2F.1 was independently predicted to be a gene (GenBank accession number 2661571) by the EMBL genefinder program subsequent to the submission of the provisional application to

10   which this application claims priority.

Homologs of the two candidates were found, and their sequences were used in turn for further database searches. In total, forty-nine genes have been identified from the approximately 16% genomic sequence currently available. Applicants have tentatively named this family of genes DOR (for *Drosophila* Olfactory Receptor), and each individual gene was

15   named based upon its cytogenetic location in the genome. Thus the two genes identified initially are DOR22A.1 and DOR22A.2, which were abbreviated here as 22A.1 and 22A.2 (the final digit in this nomenclature is used to distinguish the genes at a site and does not refer to the cytogenetic band number). The genomic locations of all the DOR genes identified so far are indicated in Figure 2A, and an alignment of their amino acid sequences is presented in

20   Figure 3. Of the forty-nine family members, the great majority have been found to be expressed in either the antenna or the maxillary palp, or in both, based upon RT-PCR analysis (Table 1) and *in situ* hybridizations to RNA in tissue sections.

The DOR genes have no significant similarities to any known genes, and do not appear in any of the *Drosophila* EST databases. However, Kyte-Doolittle hydropathy plots of

25   the predicted proteins show that each has approximately seven peaks that could represent transmembrane domains (Figure 2C) (Kyte & Doolittle, (1982) J. Mol. Biol. 157, 105-132). The lengths of the sixteen proteins are between 369 and 403 amino acids, similar to the lengths of most previously described families of GPCRs (Probst *et al.*, (1992) DNA Cell Biol.

- 56 -

57

11, 1-20). In addition, the spacing of the putative transmembrane domains gives rise to predicted intracellular and extracellular loops similar in size to those in many families of GPCRs (Probst *et al.*, (1992) DNA Cell Biol. 11, 1-20).

5   Amino acid sequence identity among the DOR genes ranges from approximately 10-75%, with many genes showing a relatively low level of identity to each other (approximately 20%). Two pairs of clustered genes, 22A.1/22A.2 and 33B.1/33B.2 show the highest identity, with 75% and 57% homology, respectively. However, not all clustered genes show high degrees of similarity. 33B.3, for example, is only 28% identical to both 33B.1 and 33B.2 and 46F.1 and 46F.2 are only 29% identical. In addition to exhibiting sequence

10   identity, many of the genes contain introns in corresponding locations (Figure 3), consistent with their constituting a family derived from a common ancestral gene. Examples of genomic DNA encoding the complete structural gene for DOR proteins containing the introns can be found in SEQ ID NO: 99-114, while the corresponding cDNA containing the intact ORF can be found in SEQ ID NO: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29 and 31.

15   There are sixty-seven residues that are conserved among at least 50% of the genes, and most of these (49) are in the C-terminal halves of the proteins (Figure 3). Among the conserved residues are a serine and a threonine in the intracellular C-terminal tail, residues frequently conserved in this region of GPCRs (Probst *et al.*, (1992) DNA Cell Biol. 11, 1-20). The most divergent region in the sequences is a stretch of thirty amino acids representing part

20   of the first extracellular loop and nearly all of transmembrane domain three. The divergence in this region also occurs in the most conserved pairs of genes: 22A.1 and 22A.2 are 75% identical overall, but only 50% identical in this region, and 33B.1 and 33B.2 are 57% identical overall, but only 33% identical in this region. This divergence has also been observed in other species. In particular, transmembrane domains three, four and five were exceptionally

25   divergent in rat odorant receptors and have been proposed to play a role in odorant binding (Buck & Axel, (1991) Cell 65, 175-187).

Some of the genes are clustered in the genome (Figure 2A), while others are apparently isolated. Within a cluster the average intergenic distance is on the order of 500

- 57 -

58

bases. Clustered DOR genes do not necessarily have introns in corresponding locations (*e.g.* 46F.1 and 46F.2), but all clustered genes have their transcriptional orientations in the same direction (Figure 2A). At least one of the DOR genes (2F.1) is flanked closely on both sides by two apparently unrelated genes (Figure 2B) (Haenlin *et al.*, (1987) EMBO J. 6, 801-807).

5   A novel strategy to search the *Drosophila* genomic sequence database for genes encoding potential GPCRs was employed, leading to the identification of a multigene family with properties expected of odorant receptors. In addition to these genes, a wide variety of other transmembrane proteins were identified by this strategy, a few previously identified by other means and many representing novel proteins with similarity to known transmembrane

10   proteins. These results suggest that the algorithm may be of widespread use in identifying new receptors, channels, and other transmembrane proteins.

The family of candidate odorant receptor genes currently contains forty-nine members, identified from the 16% of the *Drosophila* genomic sequence that is available. By extrapolation the size of this family may be on the order of 100 genes, making it the largest

15   gene family identified in *Drosophila*.

There are several lines of evidence indicating that these genes encode *Drosophila* odorant receptors. First, the predicted proteins encoded by the genes each contain approximately seven potential transmembrane domains, as expected of GPCRs. Second, genes are expressed in one or both of the two olfactory organs, and for a number of genes this

20   expression is restricted to a subset of olfactory receptors, as expected for odorant receptors. Third, the large number of family members, and the clustered location of many of these genes in the *Drosophila* genome, is reminiscent of odorant receptors in other organisms.

Additional lines of evidence is available which indicates DOR proteins as odor receptors. First, antibodies raised against the product of the DOR22A.2 gene label a small

25   number of sensilla on the fly's antenna whose location corresponds to the same region labeled by *in situ* hybridization. Most important, staining appears localized to the cavities of the labeled sensilla, where the dendritic cells are located. This is exactly the localization expected of an odorant receptor. Second, different DOR genes are expressed (as determined by *in situ*

- 58 -

hybridization) in different subsets of olfactory receptor neurons, as expected of odor receptor genes. Third, as expected, the number of olfactory receptor neurons labeled by individual DOR genes corresponds with the number of olfactory receptor neurons exhibiting a particular odor-sensitivity because the number of neurons expressing a particular DOR gene is predicted

5   to equal the number of neurons with a particular odor response spectrum. Finally, many of the DOR genes are not expressed in the Acj6 POU-domain transcription factor mutant, where a subset of olfactory receptor neurons displayed abnormal odorant specificities. A correlation between DOR gene expression and odorant-specificity therefore exists, as is expected with odorant receptor genes.

10   Comparison of the sequences of these candidate odorant receptors to those from other organisms shows that they are extremely divergent from known odorant receptors and other GPCR families. This is not surprising, as searches for these genes based on sequence similarity to odorant receptors from other organisms had not succeeded, and the odorant receptor families in vertebrates and *C. elegans* are essentially unrelated. There is a great deal

15   of sequence divergence among the DOR genes, much more than among the rat sequences previously reported (Buck & Axel, (1991) Cell 65, 175-187), for example. Moreover, genomic Southern blots have shown that none of nine DOR genes tested defines a subfamily of more than two or so well-conserved genes. The DOR family therefore differs in this respect from the mouse family, for example, where most odorant receptor genes belong to

20   subfamilies of approximately seven to ten genes (Ressler *et al.*, (1993) Cell 73, 597-609).

Although at present the clusters of DOR genes identified thus far contain smaller numbers of genes (less than three) than in other organisms (Troemel *et al.*, (1995) Cell 83, 207-218; Sullivan *et al.*, (1996) Proc. Natl. Acad. Sci. USA 93, 884-888; Barth *et al.*, (1997) Neuron 19, 359-369), a number of interesting features of the clustered genes are already

25   apparent. As found in other organisms (Barth *et al.*, (1997) Neuron 19, 359-369), *Drosophila* odorant receptor genes within a cluster are not necessarily coordinately regulated, such that genes within a cluster are expressed in different classes of cells, and even in different olfactory organs (*e.g.* 46F.1 is expressed in the maxillary palp whereas 46F.2 is expressed in the

antenna). So far, all genes identified within a cluster, however, are transcribed in the same orientation. Genes within a cluster sometimes do, but sometimes do not, share intron positions, suggesting that introns may have become lost following gene duplication; a phylogenetic study revealed extensive gene duplication and intron loss among the chemoreceptor genes of *C. elegans* (Robertson, (1998) Genome Res. 8, 449-463).

Step 3: Identification of olfactory receptor genes using RT-PCR. RT-PCR with primers designed from two of these final candidates yielded amplification products from antennal cDNA. From RT-PCR experiments, the two genes did not appear to be expressed in the maxillary palp, abdomen, thorax, or head from which olfactory organs had been removed, suggesting that these genes were expressed specifically in the antenna. These two genes are located within 500 base pairs of each other at cytological position 22A (Figure 2A), and their predicted proteins are 75% homologous at the amino acid level.

For preparation of RNA, individual flies were frozen in liquid nitrogen, and antennae and maxillary palps were dissected. On average 150 antennae or 200 maxillary palps were used for RNA preparation. Total RNA was prepared as described elsewhere (McKenna *et al.*, (1994) J. Biol. Chem. 269, 16340-16347). The RNA was treated with DNaseI (Gibco-BRL) for thirty minutes at 37°C, phenol/chloroform extracted, and precipitated. The entire RNA preparation was used for oligo dT-primed cDNA synthesis using Superscript II Reverse Transcriptase (Gibco-BRL) according to the manufacturer's directions. PCR was performed using Taq polymerase (Sigma) under standard cycling conditions, with an annealing temperature of 60°C, gene-specific primer concentration of 1 pM, and magnesium concentration of 2.5 mM. For all genes except 2F.1, primer pairs which span introns were used in order to distinguish PCR bands amplified from cDNA from those amplified from any remaining genomic DNA.

Example 2: Hybridization of DOR gene probes to related sequences

To determine whether any of the DOR genes have closely related homologs, coding regions from nine of the genes were used to probe Southern blots of *Drosophila* genomic

DNA at high or low stringency. For the closely related genes such as 22A.1 and 22A.2, a combined probe was used. For genomic southern blots, hybridizations were at 65°C (high stringency) or 55°C (low stringency), in 7% SDS, 0.5 M sodium-phosphate buffer pH 7.2, 1 mM EDTA, pH 8.0.

5   Each probe detected only its own sequence at high stringency, while at low stringency most gene probes detected one or two novel bands (data not shown). As expected, because of the overall low level of similarity, none of these extra bands corresponded to any of the other known DOR genes. These data indicate that some of these genes have one or two closely related homologs, but that none belongs to a large subfamily of highly related genes.

10

Example 3: Localization of DOR gene expression

          Olfactory receptor neurons of the adult fly are located in both the antenna and the maxillary palp. To ask whether any of the DOR genes are expressed in these neurons, *in situ* hybridization was carried out using adult tissue sections.

15   For *in situ* hybridization experiments, coding regions of the DOR genes were subcloned into the pGEM-T Easy vector (Promega). Digoxygenin-labeled RNA probes were generated and hydrolyzed according to the manufacturer's instructions (Boehringer Mannheim). *In situ* hybridizations to RNA in tissue sections were performed using a modified version of procedures described elsewhere (Roberts, (1998) *Drosophila*: A Practical

20 Approach, Oxford University Press; Chadwick & McGinnis, (1987) EMBO J. 6, 779-789). Briefly, heads were dissected from animals and fixed in 4% paraformaldehyde/PBS for fifteen minutes. Tween-20 was then added to 0.1% and heads were fixed for an additional thirty minutes. Samples were washed twice for five minutes in 0.1% Tween 20/PBS (PBST), cut into 8 μm frozen sections, and mounted on poly-L-Lysine treated slides (Sigma). Sections

25 were dried onto slides for thirty minutes at room temperature and then fixed for an additional thirty minutes in 4% paraformaldehyde/PBST. Samples were washed for a total of two hours in PBST with five changes of buffer, followed by an incubation for five minutes in 1:1 PBST:hybridization buffer (50% formamide, 5× SSC, 50 mg/ml heparin, 0.1% Tween 20),

- 61 -

62

and then prehybridized for two hours at 55°C.

Of eleven genes examined, seven displayed detectable expression, which in every case was restricted to the olfactory organs (Table 2). The 46F.1 probe hybridized to a subset of olfactory receptors in the maxillary palp (Figure 4A). Counting of labeled olfactory receptors in serial sections revealed that the total number of 46F.1-staining olfactory receptors per maxillary palp was 18±1 (Table 2), or 15% of the 120 olfactory neurons in the maxillary palp. A similar number of neurons, 17±1, was labeled by another probe, 33B.3 (Figure 4B). The neuronal identity of the labeled cells was apparent from the presence in many cases of a well-defined axon projecting from the labeled cell body and joining the maxillary nerve (Figures 4B-C). For both probes, the labeled neurons were distributed broadly over the olfactory surface of the organ, and were interspersed among unlabeled neurons (Figures 4A-C). Staining in many cells appeared annular, which was interpreted to reflect a perinuclear distribution of mRNA, as expected of an mRNA present at highest concentrations in the cell bodies of these olfactory receptors (Figure 4B). The 33B.3 and 46F.1 genes are evidently expressed in different subsets of olfactory receptors, because the number of neurons hybridizing with a mixed probe was greater than the number of neurons that hybridized when either probe was used individually (data not shown). No hybridization detected in the antenna, head, or thorax for either probe.

Many of the DOR genes are expressed in the antenna and not in the maxillary palp, as determined by RT-PCR (Table 1). For several genes this localization was confirmed by *in situ* hybridization. The 47E.1 probe hybridized to 40±1 cells in a broad area across the antenna (Figures 5A-B), including both anterior and posterior faces, similar to the distribution pattern of small s. basiconica (Figure 1F). A probe from the 25A.1 gene hybridized to fewer cells, 16±1, but in a region of the antenna similar to that of 47E.1 staining, as judged by reconstruction of serial sections (Figure 5C-D). The 22A.2 probe hybridized to 22±1 cells in a different distribution, clustered in the dorso-medial region of the antenna (Figure 5E). This pattern matches the distribution of the large s. basiconica (Figure 1E). The expression patterns of the three genes in the antenna are illustrated schematically in Figure 5G. None of

63

these three probes revealed expression in the maxillary palp, head, or thorax. This data demonstrates that the DOR family is expressed in olfactory receptors, and that the expression of individual members is restricted to distinct subsets of cells in the olfactory organs.

5   The number and broad distribution of maxillary palp neurons expressing 46F.1 and 33B.3 are intriguing in light of electrophysiological studies. There are approximately 120 olfactory receptors on the palp, which fall into six different classes based upon their odorant response profiles. Each class contains roughly equal numbers of neurons, distributed broadly over the olfactory surface of the palp. Thus, if an individual receptor gene is expressed in all olfactory receptors of a functional class, one might expect a gene to be expressed in a broad

10   distribution, in approximately twenty neurons, in good agreement with the distribution and numbers observed for both 46F.1 and 33B.3 (18±1 and 17±1, respectively).

The two DOR genes whose expression was detected by *in situ* hybridization in the maxillary palp are expressed in olfactory receptors housed within s. basiconica, the only morphological class of sensilla on the palp. In the antenna, the 22A.2 probe consistently

15   hybridized to a subset of cells in a portion of the dorso-medial region of the antenna that contains almost exclusively large s. basiconica (Figure 1E). The 47E.1 and 25A.1 probes hybridize to subsets of cells in a distinctly different region of the antenna which may correlate with the distribution of small s. basiconica, of which at least two functional types are intermingled (Figure 1F). Of particular interest, the numbers of cells to which 47E.1 and

20   25A.1 hybridize are different: 40±1 and 16±1; one possible interpretation is that they are expressed in distinct functional types of small s. basiconica. This region also contains s. trichodea and s. coeloconica, and although the labeling patterns do not correlate with the distribution of either of two functional classes of s. trichodea (Clyne *et al.*, (1997) Invert. Neurosci. 3, 127-135), a definitive identification of the sensillar type may require further

25   investigation. If in fact all the DOR genes are expressed in only one of the morphological categories of sensilla, the s. basiconica, it is possible that there are other, as yet unidentified, families of receptors that are expressed in the other morphological categories of sensilla. This would mean that the number of odorant receptors in *Drosophila* might be substantially larger

than one-hundred.

Applicants have identified three DOR genes that are expressed in the maxillary palp (Table 1), from the 16% of the genome analyzed. As these three genes, like most DOR genes, are not clustered in the genome, linear extrapolation suggests that the entire genome contains

5   on the order of eighteen DOR genes expressed in the maxillary palp, an organ which has six functional classes of neurons (Clyne *et al.*, (1999) Neuron 22, 339-347; de Bruyne *et al.*, (1999) J. Neurosci. 19, 4520-4532). If all neurons within a functional class, *i.e.* with the same odor-specificity, are identical in terms of their receptor expression, then the ratio of expressed genes to neuronal classes in this organ would be consistent with a model in which an

10   individual ORN expresses a small number of odorant receptors; however, further data is needed to establish conclusively the number of receptor genes expressed per cell. Olfactory neurons in other organisms appear to lie at either of two extremes: in the vertebrates, it is believed only one receptor is expressed per ORN (Ngai *et al.*, (1993) Cell 72, 667-680; Ressler *et al.*, (1993) Cell 73, 597-609; Vassar *et al.*, (1993) Cell 74, 309-318); in *C. elegans*,

15   approximately 550 chemoreceptors are likely to be distributed amongst fourteen classes of chemosensory neurons (Troemel *et al.*, (1995) Cell 83, 207-218).

Olfactory receptors in *Drosophila* and other insects project to an olfactory processing center, the antennal lobe, which is much like the olfactory bulb of vertebrates. Like its vertebrate counterpart, the antennal lobe contains olfactory glomeruli, of which the antennal

20   lobe of *Drosophila* has approximately forty (Stocker *et al.*, (1995) Roux's Arch Dev Biol 205, 62-72; Laissue *et al.*, (1999) J. Comp. Neurol. 405, 543-552). In vertebrates there is an approximate equivalence between the estimated number of odorant receptor genes and the number of glomeruli (Barth *et al.*, (1996) Neuron 16, 23-34; Buck, (1996) Annu. Rev. Neurosci. 19, 517-544); since *C. elegans* does not contain glomeruli, it has not been possible

25   until now to consider whether the evolutionary conservation of this equivalence extends to invertebrates. If in fact the number of DOR genes is one-hundred, then the ratio of odorant receptor genes to glomeruli would exceed two, and would rise if additional families of odorant receptor genes were discovered. Of particular interest, the number of glomeruli receiving

- 64 -

input from the maxillary palp has been variously estimated as three and five (Venkatesh & Singh, (1984) Int. J. Insect. Morphol. Embryol. 13, 51-63; Stocker *et al.*, (1995) Roux's Arch Dev Biol 205, 62-72); if our estimate of eighteen genes expressed in the maxillary palp is correct, then the ratio of these receptor genes to their corresponding glomeruli would fall in

5   the range of three to six.

Example 4: DOR gene expression during development

Recent evidence supports a dual role for the vertebrate olfactory receptors. First, these receptors have an instructive role in guiding the axons of olfactory receptors to the correct

10   glomeruli during development (Mombaerts *et al.*, (1996) Cell 87, 675-686; Wang *et al.*, (1998) Cell 93, 47-60), and second as odorant receptors in the adult (Zhao *et al.*, (1998) Science 279, 237-242). To address the possibility that the DOR genes might also play a role in development, three DOR probes were hybridized to antennal sections from different stages of pupal development. In *Drosophila*, ORN axons first leave the developing antenna at

15   approximately sixteen hours after puparium formation (APF) (Lienhard & Stocker, (1991) Development 112, 1063-1075; Ray & Rodrigues, (1995) Dev. Biol. 167, 426-438; Reddy *et al.*, (1997) Development 124, 703-712), and the diameter of the antennal nerve continues to increase until 72 hours APF (Stocker *et al.*, (1995) Roux's Arch. Dev. Biol. 205, 62-72). Glomeruli first become visible in the antennal lobe at approximately 48 hours APF.

20   Developing antennae were therefore examined at 16, 24, 36, 48, 54, 60, 72 and 93 hours APF (adults eclosed from the pupal case at approximately 100 hours). For these developmental studies, *Drosophila* were collected as white prepupae and kept at 25°C on moist filter paper for the indicated number of hours, at which time they were fixed. At 25°C the approximate time from the white prepupal stage to eclosion is 100 hours (Lockett & Ashburner, (1989)

25   Dev. Biol. 134, 430-437).

Cells positive for 22A.2 were first seen at 60 hours APF, indicating that detectable expression begins between 54 and 60 hours, well within the period in which the antennal nerve is still increasing in diameter (Figure 6A-B). A subset of cells was labeled at this time,

and they were restricted to a subregion of the developing antenna; the pattern appears comparable to that of the mature antenna, although this pattern was not characterized in as much detail as that of the adult. Labeling with 22A.2 was also observed in antennae at all subsequent time points. Interestingly, cells positive for 47E.1 and 25A.1 were not observed

5　until much later, at the 93 hour time point; they were not observed at any of the earlier times (Figure 6C-D and data not shown). For comparison, *in situ* hybridization was also performed with a probe representing the odorant-binding protein OS-E (McKenna *et al.*, (1994) J. Biol. Chem. 269, 16340-16347), which is believed to play a role in olfactory function, but which has not been implicated in a developmental process. OS-E was also first observed at 93 hours,

10　at which time it expression increased (Figure 6E-F).

Example 5: Regulation of DOR expression by POU domain transcription factor *acj6*

　　　　Little is known about the regulation of odor receptor genes, a process critical to the establishment of olfactory neuron identity and ultimately to the process of olfactory coding.

15　In *C. elegans* the *odr7* gene, a member of the nuclear receptor superfamily, has been shown to regulate the odorant receptor gene *odr10* (Sengupta *et al.*, (1994) Cell 79, 971-980; Sengupta *et al.*, (1996) Cell 84, 899-909). In *Drosophila*, null mutations of the *acj6* gene, which encodes a POU domain transcription factor, eliminate the odor response of three of the six classes of maxillary palp olfactory receptors (Clyne *et al.*, (1999) Neuron 22, 339-347). A

20　fourth ORN class on the maxillary palp is altered to a new class of ORN with a novel odor sensitivity. These data suggest that Acj6 plays a role in the differentiation of certain maxillary palp olfactory receptors, perhaps by determining which olfactory receptor gene(s) are expressed. To address the possibility that Acj6 regulates odorant receptor genes, probes from the 33B.3 and 46F.1 genes were hybridized to sections of maxillary palps from the null

25　mutant, *acj6[6]*. No hybridization was detected in either case (Figure 4D and data not shown), nor was expression of either gene detected by RT-PCR from *acj6[6]* maxillary palps (Table 1).

　　　　acj6 mutations also affect the physiological response of the antennal neurons to odors (Ayer & Carlson, (1991) Proc. Nat. Acad. Sci. USA 88, 5467-5471; Ayer & Carlson, (1992)

J. Neurobiol. 23, 965-982). 22A.2, 25A.1, and 47E.1 probes were therefore hybridized to sections of *acj6⁶* antennae. All three probes hybridized to groups of cells in the same locations as in the wild type antenna (Figure 5F and data not shown). RT-PCR amplification showed that expression of certain other DOR genes, 33B.1, 33B.2, 33B.3, and 46F.2 was eliminated

5   in the antenna of *acj6⁶* (Table 1). Thus, in the *acj6⁶* mutant, one subset of candidate odorant receptor genes was not expressed while a different subset remained unaffected. Interestingly, genes within a cluster all showed similar dependency on Acj6: 33B.1, 33B.2, and 33B.3, for example, all depended on Acj6, whereas 22A.1 and 22A.2 did not. In summary, these data support a role for *acj6* in the regulation of a subset of olfactory receptor genes.

10   The DOR family is subject to complex regulation. First, the expression of individual DOR genes exhibits highly specific tissue and spatial localization. Some genes are expressed in the antenna but not the maxillary palp; others show expression in the maxillary palp but not the antenna. Within an organ, expression of a particular DOR gene is restricted to a subset of cells. In the antenna, the patterns of expression are spatially regulated, exhibiting regional

15   specificity of expression as detailed above. In the maxillary palp, expression is limited to a population of neurons approximately equal in number to the neurons of a functional class.

DOR genes are also subject to interesting temporal regulation. One gene, 22A.2, is expressed in the developing antenna during a time when the antennal nerve is still increasing in diameter (Stocker *et al.*, (1995) Roux's Arch. Dev. Biol. 205, 62-72) . These data leave

20   open a possible role for *Drosophila* olfactory receptors in axon guidance and glomerulus formation, a role for which evidence has been found in vertebrates (Mombaerts *et al.*, (1996) Cell 87, 675-686; Wang *et al.*, (1998) Cell 93, 47-60) but not *C. elegans*. In zebrafish, odorant receptors show asynchronous onset of expression during development of the olfactory placode (Barth *et al.*, (1996) Neuron 16, 23-34). The DOR genes also show heterogeneity in

25   their temporal regulation: expression of two other DOR genes begins much later than for the 22A.2 gene. If in fact individual olfactory receptors express more than one DOR gene, perhaps some have acquired a specialized role in development.

Evidence also exists indicating that different DOR genes are expressed at different

levels of abundance within cells. Although RT-PCR experiments demonstrated expression of 25A.1 in both antenna and maxillary palp, *in situ* hybridization revealed expression of 25A.1 only in the antenna of each animal examined; conversely, although RT-PCR experiments showed expression of 33B.3 in both olfactory organs, *in situ* hybridization detected label only

5   in the maxillary palp of each animal examined (Tables 1 and 2). These results suggest that a receptor gene may be expressed at different cellular levels in the two organs, and that different genes may be expressed at different cellular levels in the same organ. Such an explanation would suggest that there are mechanisms governing not only the spatial and temporal control of DOR genes, but also their levels of expression.

10   If DOR genes are in fact expressed at different cellular levels in particular olfactory receptors, then perhaps the four DOR genes that were undetectable in the antenna by *in situ* hybridization, despite clear evidence for their antennal expression from RT-PCR, a more sensitive technique, are among those expressed at low levels. It is important to note that in *C. elegans*, expression of a number of candidate odorant receptors was undetectable using GFP

15   fusion genes (Troemel *et al.*, (1995) Cell 83, 207-218).

As a first step in investigating the mechanisms through which the complex regulation of DOR genes is achieved, the role of the POU domain transcription factor Acj6 was tested, which was previously found to act in governing olfactory neuron identity. Applicants found that Acj6 is in fact required for expression of the DOR family. Two lines of evidence,

20   RT-PCR and *in situ* hybridization analysis, both indicate that proper expression of a specific subset of DOR genes depends on Acj6. The results indicate that the odor-specificity of a subset of olfactory receptors is governed at least in part by the action of the Acj6 POU domain transcription factor on DOR genes, and are fully consistent with the notion that DOR genes encode odorant receptors.

25   The isolation of genes likely to encode odorant receptors in *Drosophila* opens a number of avenues for future investigation. Drosophila provides the ability to manipulate odor receptors genetically and test the functional consequences of such manipulations *in vivo*, either physiologically or behaviorally. Such analysis may be useful in examining potential

- 68 -

69

roles of DOR proteins in olfactory response and in development. It may also be possible to isolate homologous genes in other insects, including some which provide excellent opportunities for research and some of agricultural or medical importance which rely on olfactory cues to locate their hosts.

5

Example 6: Transgenic *Drosophila*

P element mediated germline transformation of Drosophila can be carried out as previously described (Rubin & Spradling, (1982) Science 218, 348-353). *Drosophila* embryos are isolated and microinjected with P element expression constructs as previously

10   described (Karess & Rubin, (1984) Cell 38, 135-146) containing a particular DOR nucleotide sequence, at 0.5 mg/ml together with a helper plasmid at 0.1 mg/ml. G_0 injected adults are individually back crossed to the recipient strain and the G_1 progeny screened for the w+ transformation marker (Klemenz *et al.*, (1987) Nucleic Acids Res. 10, 3947-3959). Transformed lines homozygous for the transgene are established from orange eyed G_1 flies as

15   previously described (Klemenz *et al.*, (1987) Nucleic Acids Res. 10, 3947-3959).

A line of *Drosophila* in which the DOR33B.3 gene can be over-expressed was constructed as described above. The DOR33B.3 coding sequences were joined to an upstream activating sequence (UAS) and introduced by P element-mediated germline transformation into *Drosophila*. A yeast GAL4 transcription factor gene, coupled to a heat

20   shock promoter, was then crossed into the transgenic line. As expected, heat shock of this line resulted in induction of DOR33B.3 expression. The heat shock-induced expression of GAL4, results in binding of GAL4 to the UAS, and subsequent induction of DOR33B.3 expression. This transgenic line of *Drosophila*, and three other transgenic lines containing other DOR genes, can be tested for elevated responses to any of fifty different odors. Elevated response

25   to any particular odorant is indicative of an ligand which binds and activates the over-expressed receptor (see, *e.g.*, Zhao & Firestein, (1998) Science 279, 237-242).

Although the present invention has been described in detail with reference to

examples above, it is understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims. All cited patents and publications referred to in this application are herein incorporated by reference in their entirety. The results of the experiments disclosed herein

5   have been published in the journal Neuron (22, 327-338) in February, 1999, this article herein incorporated by reference in its entirety.

71